CENTRO INTERNACIONAL DE ESTUDOS DE DOUTORAMENTO E AVANZADOS DA USC (CIEDUS)

TESIS DE DOCTORADO

# ENSEMBLES OF CHOICE-BASED MODELS FOR RECOMMENDER SYSTEMS

**Ameed Ali Almomani**

**ESCUELA DE DOCTORADO INTERNACIONAL**

**PROGRAMA DE DOCTORADO EN Investigación en Tecnoloxías da Información**

SANTIAGO DE COMPOSTELA

2020

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

**DECLARACIÓN DO AUTOR DA TESE ENSEMBLES OF CHOICE-BASED MODELS FOR RECOMMENDER SYSTEMS**

D. AMEED ALI ALMOMANI

Presento a miña tese, seguindo o procedemento axeitado ao Regulamento, e declaro que:

1. A tese abarca os resultados da elaboración do meu traballo.
2. De selo caso, na tese faise referencia ás colaboracións que tivo este traballo.
3. A tese é a versión definitiva presentada para a súa defensa e coincide coa versión enviada en formato electrónico.
4. Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.

En SANTIAGO DE COMPOSTELA, 18 de Junio de 2020

Asdo. AMEED ALI AHMAD ALMOMANI

**AUTORIZACIÓN DOS DIRECTORES DA TESE: ENSEMBLES OF CHOICE-BASED MODELS FOR RECOMMENDER SYSTEMS**

Dr. Eduardo Sánchez Vila.

**INFORMAN**:

Que a presente tese, correspóndese co traballo realizado por **D. AMEED ALI ALMOMANI**, baixo a miña dirección, e autorizo a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como director desta non incorre nas causas de abstención establecidas na Lei 40/2015.

En SANTIAGO DE COMPOSTELA, 18 de Junio de 2020

Asdo. Eduardo Sánchez Vila

**I dedicate this work**

***For soul my Dad, Who has inspired me that nothing is impossible with working hard.***

***For my Mom, Wife, and Kids, who are the most beautiful thing in my life.***

*"The truth may be puzzling. It may take some work to grapple with. It may be counterintuitive. It may contradict deeply held prejudices. It may not be consonant with what we desperately want to be true. But our preferences do not determine what's true."*
*Carl Sagan*

## Acknowledgments

In the beginning, the Ph.D. has been a truly life-changing experience for me where I learned and acquired different skills, and taken advantage of unique opportunities in my life. This includes participating in important events and collaborations derived from the Ph.D. It would not have been possible to do without the support and guidance that I received from remarkable people.

I would like to express my special thank and appreciation to my supervisor Prof. Eduardo Sanchez Vila. I would like to thank you for assisting my research and allowing me to grow and improving research skills. Thanks for your patience, motivation, and extensive knowledge, as well as, for the endless support of my Ph.D research. Your supervision assisted me in the entire research and writing of this thesis.

Thanks and great appreciation to my wife, who helped me and took responsibility during my studies. My mother, your prayer for me was what helped me thus far. I would like to thanks my brothers, and sisters for supports in this stage of my life.

2020

## Resumo

Nesta tese centrámonos en tres principais paradigmas: Sistemas de recomendadores, toma de decisións e ensembles. Para este propósito, este traballo estrutúrase do seguinte xeito.

En primeiro lugar, a tese analiza o potencial de modelos baseados na elección. A motivación detrás baseouse na idea de aplicar paradigmas de decisión sólidos, como a teoría da elección e a utilidade, no campo dos Sistemas de Recomendadores. Desde esta perspectiva, considérase que o problema de recomendación é un problema de predición de elección que de predición de valoración. A motivación deste traballo foi superar algunhas das limitacións dos algoritmos actuais de recomendadores de última xeración proporcionando: (1) preferencias precisas, que se aprenden das opcións do usuario en lugar das clasificacións e (2) a transparencia, que se consegue. mediante o conxunto de coeficientes estimados dos modelos de elección.

En segundo lugar, esta investigación analiza o proceso cognitivo subxacente ao comportamento de elección. Por un lado, a actividade neuronal e a mirada rexistráronse experimentalmente de diferentes temas realizando unha tarefa de elección nunha Interfaz Web. Doutra banda, os cognitivos encaixáronse empregando trazos racionais, emocionais e de atención. Avaliaban as previsións do modelo en función da súa precisión e fixéronse os rankings para cada usuario. Os resultados mostran que: (1) os modelos atencionais son os mellores en termos do seu rendemento medio en todos os usuarios (2) cada tema amosa un mellor modelo diferente, e (3) os conxuntos poden funcionar mellor que os dunha soa opción, pero un óptimo. hai que atopar un método de construción.

Finalmente, o traballo explora a hibridación de modelos baseados na elección con conxuntos. O obxectivo é aproveitar o mellor dos dous mundos: transparencia e rendemento. Analizáronse dous métodos principais para construír conxuntos baseados na elección óptimos: desinformados e informados. Para o primeiro, avaliáronse dúas estratexias: os grupos de 1-Learners e N-Learners. Os resultados sinalan o rendemento superior dos grupos de N-Learners e amosan o potencial de arranxos personalizados. Para a segunda, confiamos en tres tipos de información previa: (1) Alta diversidade, (2) Predición de erro baixo (MSE), (3) e erro de multitude baixa. Empregouse un enfoque Greedy para seleccionar os modelos individuais e para construír os conxuntos. De novo, analizáronse dúas estratexias, modelos comúns e personalizados en canto á súa precisión así como a frecuencia do mellor modelo. Os resultados mostran o rendemento superior dos conxuntos informados e indican a baixa MSE como a clave antes de construír conxuntos baseados na elección óptima.

**Palabras chave**: Conxuntos, Árbores de decisións, Recomendacións, Sistemas de recomendadores, Modelos de elección, Toma de decisións, Modelos de elección Ensembles, Teoría do proceso dobre.

# Resumen


En esta tesis, nos centramos en tres paradigmas principales: sistemas de recomendación, toma de decisiones y conjuntos. Para este propósito, este trabajo se estructura de la siguiente manera.

Primero, la tesis analiza el potencial de los modelos basados en la elección. La motivación detrás de esto se basó en la idea de aplicar paradigmas sólidos de toma de decisiones, como la elección y la teoría de la utilidad, en el campo de los sistemas de recomendación. Desde esta perspectiva, el problema de recomendación se considera un problema de predicción de elección más que de predicción de calificación. La motivación detrás de este trabajo fue superar algunas de las limitaciones de los algoritmos de recomendación de última generación al proporcionar: (1) preferencias precisas, que se aprenden de las elecciones del usuario en lugar de las calificaciones, y (2) transparencia, que se logra mediante el conjunto de coeficientes estimados de los modelos de elección.

En segundo lugar, esta investigación analiza el proceso cognitivo subyacente al comportamiento de elección. Por un lado, la actividad neuronal y la mirada se registraron experimentalmente de diferentes sujetos realizando una tarea de elección en una interfaz web. Por otro lado, los cognitivos se ajustaron utilizando características racionales, emocionales y atencionales. Las predicciones del modelo se evaluaron en términos de su precisión y se hicieron clasificaciones para cada usuario. Los resultados muestran que: (1) los modelos atencionales son los mejores en términos de rendimiento promedio para todos los usuarios, (2) cada sujeto muestra un mejor modelo diferente, y (3) los conjuntos pueden funcionar mejor que los modelos de opción única pero un óptimo El método de construcción tiene que ser encontrado.

Finalmente, el trabajo explora la hibridación de modelos basados en elecciones con conjuntos. El objetivo es tomar lo mejor de los dos mundos: transparencia y rendimiento. Se analizaron dos métodos principales para construir conjuntos óptimos basados en la elección: desinformado e informado. Para el primero, se evaluaron dos estrategias: conjuntos de 1 alumno y N-alumnos. Los resultados señalan el rendimiento superior de los conjuntos de N-Learners y muestran el potencial de los arreglos personalizados. Para el segundo, confiamos en tres tipos de información previa: (1) Alta diversidad, (2) Predicción de error bajo (MSE),

(3) y Error de multitud baja. Se usó un enfoque codicioso para seleccionar los modelos individuales y construir los conjuntos. Una vez más, se analizaron dos estrategias, modelos comunes y personalizados en términos de su precisión, así como la frecuencia del mejor modelo. Los resultados muestran el rendimiento superior de los conjuntos informados e indican el bajo MSE como la clave antes de construir conjuntos óptimos basados en la elección.

**Palabras clave**: Conjuntos, árboles de decisión, recomendaciones, sistemas de recomendación, modelos de elección, toma de decisiones, conjuntos de modelos de elección, teoría del proceso dual.

## Summary

In this thesis, we focused on three main paradigms: Recommender Systems, Decision Making, and Ensembles. For this purpose, this work is structured as follows.

First, the thesis analyzes the potential of choice-based models. The motivation behind this was based on the idea of applying sound decision-making paradigms, such as choice and utility theory, in the field of Recommender Systems. From this perspective, the recommendation problem is considered a problem of choice prediction rather than rating prediction. The motivation behind this work was to overcome some of the limitations of current state-of-art recommender algorithms by providing: (1) accurate preferences, which are learnt from user choices rather than from ratings, and (2) transparency, which is achieved by means of the set of estimated coefficients of the choice models.

Second, this research analyzes the cognitive process underlying choice behavior. On the one hand, neural and gaze activity were recorded experimentally from different subjects performing a choice task in a Web Interface. On the other hand, cognitive were fitted using rational, emotional, and attentional features. The model's predictions were evaluated in terms of their accuracy and rankings were made for each user. The results show that: (1) the attentional models are the best in terms of its average performance across all users, (2) each subject shows a different best model, and (3) ensembles may perform better than single choice models but an optimal building method has to be found.

Finally, the work explores the hybridization of choice-based models with ensembles. The goal is to take the best of the two worlds: transparency and performance. Two main methods were analyzed to build optimal choice-based ensembles: uninformed and informed. For the first one, two strategies were evaluated: 1-Learner and N-Learners ensembles. The results point out the superior performance of N-Learners ensembles and show the potential of personalized arrangements. For the second one, we relied on three types of prior information: (1) High diversity, (2) Low error prediction (MSE), (3) and Low crowd error. A Greedy approach was used to select the single models and to construct the ensembles. Again, two strategies, common and personalized models were analyzed in terms of their accuracy as well as the frequency of the best model. The results show the superior performance of informed ensembles, and indicate the low MSE as the key prior to build optimal choice-based ensembles.

**Keywords**:Ensembles, Decision trees, Recommendations, Recommender Systems, Choice models, Decision-making, Choice models Ensembles , Dual Process Theory.

# Resumen extendido

Los sistemas de recomendación (RS) son herramientas y métodos de software que presentan sugerencias para que los artículos tengan valor para un usuario [36]. Para ello, los RS reúnen información sobre las preferencias de los usuarios por un conjunto de artículos, que pueden obtenerse explícita o implícitamente. También pueden utilizar las características demográficas de los usuarios [9]. En otro sentido, la tarea de las RS es convertir los datos sobre los usuarios y sus preferencias en predicciones de los posibles gustos e intereses futuros de los usuarios [30]. Las propuestas que proporciona un SR tienen por objeto apoyar a sus usuarios en diversos procesos de adopción de decisiones, además de ser un medio valioso para hacer frente a la sobrecarga de información. Se han sugerido varias técnicas para la generación de recomendaciones.

El desarrollo de las RS es un esfuerzo multidisciplinario que incluye especialistas de diferentes campos, como la inteligencia artificial, la minería de datos, la estadística, los sistemas de apoyo a las decisiones, el marketing, la interacción entre el hombre y la computadora, el comportamiento de los consumidores y la psicología [36]. Como resultado, se pueden utilizar en diferentes campos, y al recopilar la información disponible en la web, la recomendación del usuario se convierte en personalizada. Para ello, se construye una representación interna del usuario basada en los datos recogidos sobre sus características como edad, sexo, intereses, preferencias, etc. Pueden existir varios niveles de personalización dentro de un sistema utilizando características explícitas del usuario (por ejemplo, edad, sexo, demografía) o un patrón implícito de comportamiento del usuario (historial de navegación en la web, modelo de clic). De los contenidos generados por los usuarios pueden deducirse muchas variables ocultas, como las personalidades, las emociones y los estados de ánimo, que normalmente no son proporcionados explícitamente por los usuarios [41].

Esta nueva generación de recomendadores "humanos", también llamados sistemas de recomendación basados en la personalidad [42, 21, 15, 33], deberían considerar el "carácter" de los usuarios como algo más que un simple comportamiento individual, sino un grupo extenso y complejo de ideas, sentimientos y comportamientos.

El problema de las RS puede abordarse de diferentes maneras, considerándolo como el problema de la predicción de las elecciones del usuario en cualquier contexto particular. Bajo este punto de vista, la Teoría de la Elección Racional puede considerarse el modelo clásico utilizado para explicar las elecciones realizadas por los agentes racionales [40]. Para poder predecir las elecciones de los usuarios, necesitamos entender el proceso de toma de decisiones de los humanos. El estado del arte sobre la DM está controlado por la Teoría de Procesos Duales que considera dos tipos principales de procesos cognitivos [46]: El primero es rápido, inconsciente y sin esfuerzo; el segundo es lento, consciente y consumidor de energía. Estos procesos fueron asociados a los sistemas cognitivos por el psicólogo Daniel Kahneman, quien acuñó los términos sistema I y sistema II [25].

Por lo tanto, para construir recomendadores aún más exitosos, debemos combinar los diseños técnicos para las aplicaciones de los recomendadores con un profundo conocimiento de los procesos de toma de decisiones de los humanos [23]. Las siguientes motivaciones surgieron durante un análisis del estado de la técnica:

1. El hecho de que cada vez es más necesario explicar por qué un sitio web recomienda algunos artículos y no otros.

2. Superar algunas de las limitaciones de los actuales algoritmos de recomendación de vanguardia proporcionando: 1) preferencias precisas, que se aprenden de las elecciones de los usuarios más que de las clasificaciones, y 2) transparencia.

3. Analiza el proceso de toma de decisiones que subyace al comportamiento de elección. Primero, la actividad neuronal y de la mirada se registraron experimentalmente de diferentes sujetos que realizaban una tarea de elección en una Interfaz Web. Segundo, los modelos de elección se ajustaron usando características racionales, emocionales y de atención.

4. Nuestro estudio fue probar esta línea de pensamiento desarrollando un experimento de elección y registrando un número de variables relevantes que pudieran explicar el comportamiento observado.

5. Encontrar el método óptimo para construir conjuntos usando agregaciones de modelos de elección única.

Como se ha descrito anteriormente, el objetivo principal de esta tesis es PRESENTAR NUEVAS TÉCNICAS Y MÉTODOS PARA DESARROLLAR ALGORITOS DE SISTEMAS RECOMENDADOS basados en el uso de los modelos basados en la elección y aplicar sus modelos de conjuntos para proporcionar más precisión y explicar el comportamiento de los responsables de la toma de decisiones. Para probar nuestros sistemas, se utilizaron dos conjuntos de datos, primero, el conjunto de datos Gastronómico, que se centra en el dominio del turismo, y segundo, el conjunto de datos MOVISTAR, que se centra en el dominio de la película y el entretenimiento. Los objetivos específicos de la tesis son los siguientes:

1. Explorar el potencial y analizar el desempeño de los modelos basados en la elección. El concepto se toma prestado de la elección y la teoría de la utilidad, en la que las elecciones del usuario son los datos clave que se utilizan para conocer las preferencias de quienes toman las decisiones.

2. Desarrollar modelos cognitivos sobre la base de modelos generales basados en la elección. Al recurrir a la teoría del proceso dual, hemos creado modelos cognitivos del sistema I y del sistema II.

3. Proponer la hibridación de conjuntos de alto rendimiento con modelos transparentes e interpretables basados en la elección. Específicamente, nos hemos preguntado acerca de los métodos óptimos para construir conjuntos basados en elecciones. Más específicamente, analizamos métodos no informados o ciegos, así como informados, para seleccionar el conjunto de alumnos en el conjunto de la manera más eficiente.

En esta tesis doctoral, abordamos el objetivo primario descomponiéndolo en tres objetivos que abarcan varios aspectos de los conjuntos basados en la elección para una recomendación.

Para lograr estos objetivos, se formulan un conjunto de hipótesis de investigación relacionadas con los objetivos.

*Las primeras preguntas se refieren al primer objetivo específico. Provienen de dos observaciones en el dominio de los sistemas de recomendación: (1) los modelos basados en la calificación suponen una fuerte relación entre las preferencias y las calificaciones, (2) y las personas se muestran reacias a proporcionar recomendaciones explícitas, ya que las perciben como el resultado de un modelo publicitario Utilizamos modelos basados en la elección para superar estos problemas.*

**Q1:** ¿Qué es más relevante para generar recomendaciones: calificaciones o elecciones? **H1.** Las elecciones de los usuarios son los datos clave que se utilizan para conocer las preferencias de los responsables de la toma de decisiones. **Q2:** ¿Cuál es el rendimiento de los modelos basados en elecciones comparados con los algoritmos basados en calificaciones y conjuntos básicos? **H2.** Los algoritmos basados en la elección pueden superar el rendimiento de los algoritmos y conjuntos básicos.

*El siguiente conjunto de preguntas se refiere al segundo objetivo específico. Hemos desarrollado tres tipos diferentes de modelos cognitivos: emocional, atencional y racional. El último puede clasificarse como un modelo puro del Sistema II, mientras que el primero y el segundo podrían ser convertidos en modelos del Sistema I. Luego planteamos algunas preguntas sobre su desempeño*

**Q3:** ¿Cuál es el modelo más adecuado? S-I-emocional, S-I-atencional, o S-II-racional? **H3.**: Basándonos en la evidencia de los campos de la economía del comportamiento y el neuromercado, creemos que los modelos S-I se desempeñan mejor que los modelos S-II. Más concretamente, considerando la hipótesis del marcador somático [13], esperamos que el modelo S-I-emocional sea el mejor candidato para predecir las elecciones humanas. **Q4:** ¿Tendría sentido hablar del mejor modelo para todos los usuarios? **H4.**: Sí. Siguiendo el mismo argumento del punto anterior, creemos que los modelos emocionales S-I serán los mejores predictores para todos los usuarios.

**Q5:** ¿Podría superarse el rendimiento del mejor modelo de elección individual usando un conjunto de estudiantes racionales-emocionales-atentos?**H5**: Si. Según el estado de la técnica en el aprendizaje automático, los conjuntos suelen proporcionar el mejor rendimiento aprovechando la agregación de la información de diferentes alumnos [34, 27]. **Q6:** ¿Sería correcto hablar del mejor conjunto para todos los usuarios? **H6:** Sí. Se espera que los conjuntos construidos con características S-I-emocionales sean los mejores entre todos los usuarios.

*Luego, nos enfocamos en el tercer objetivo específico, y aplicamos métodos no informados o ciegos para construir conjuntos basados en la elección. Por lo tanto, se asumió que no hay información previa disponible sobre el desempeño de los alumnos del conjunto. Las preguntas e hipótesis específicas que han guiado esta parte del trabajo se describen brevemente:*

**Q7:** ¿Qué método es el mejor para construir conjuntos ciegos basados en la elección? Se han estudiado dos estrategias: los conjuntos de tipo 1-Learner y N-Learners. **H7:** Considerando la diversidad como un factor clave del rendimiento del conjunto, el esquema de N-aprendices debería proporcionar los mejores resultados. **Q8:** ¿Los conjuntos basados en la elección superarán a los modelos de elección única? **H8**: Esperamos que los mejores conjuntos basados en la elección funcionen mejor que los correspondientes modelos de elección única. **Q9:** ¿Qué tan buenos son los conjuntos basados en la elección comparados con los conjuntos construidos con otros modelos? Como modelo base, hemos recurrido a los Árboles de Decisión (DT), que proporcionan transparencia e interoperabilidad. **H9:** Predecimos que el conjunto ciego basado en la elección mostrará un mejor rendimiento que los conjuntos basados en DT.

*Finalmente, completamos el estudio sobre el tercer objetivo explorando métodos informados para construir conjuntos basados en la elección. Por lo tanto, se asumió que se dispone de información previa sobre el desempeño de los alumnos del conjunto. Las preguntas e hipótesis específicas que han guiado este trabajo se describen brevemente:*

**Q10:** ¿Cuál es el mejor método para construir conjuntos basados en una elección informada? Tres tipos de información previa han sido los estudios Diversidad, MSE, y Error de la Multitud. **H10:** Considerando la diversidad como información previa esencial del desempeño de los conjuntos, la diversidad debería proporcionar el mejor resultado. **Q11:** ¿El método de los conjuntos basados en la elección informada superará al modelo de elección única? **H11:** Esperábamos que los conjuntos basados en la elección mejor informados funcionaran mejor que los correspondientes modelos únicos. **Q12:** ¿Qué tan buenos son los conjuntos basados en la elección comparados con los conjuntos construidos con otros modelos? Como modelo base, hemos recurrido a los árboles de decisión (DT), que proporcionan transparencia e interpretabilidad. **H12:** Predecimos que un conjunto basado en la elección informada mostrará un mejor rendimiento que los conjuntos basados en DT. **Q13:** ¿Qué método es el mejor para construir conjuntos óptimos basados en la elección? Se han estudiado dos métodos: Ciego en conjuntos desinformados e informados. **H13:** esperábamos que el conjunto informado mostrara un mejor rendimiento que los conjuntos ciegos.

Para alcanzar estos objetivos, se ha requerido el uso de diferentes métodos de investigación, adecuados a los experimentos y estudios desarrollados para este propósito. La metodología de investigación presentada en esta tesis doctoral se desarrolló en varios experimentos.

Este doctorado La tesis realiza diversas contribuciones científicas que nos permitieron alcanzar el objetivo principal y podrían facilitar un aumento en el uso de modelos basados en la elección en los sistemas de recomendación. Estas contribuciones son las siguientes:

- Primero, el uso de un enfoque de modelo basado en la elección puede resolver el punto de la precisión de las preferencias aprendidas de las calificaciones. Si bien las calificaciones describen un resultado posterior a la experiencia que puede depender más de la satisfacción de las expectativas de los usuarios anteriores que de las preferencias, las elecciones son el resultado de la correspondencia directa entre las preferencias del usuario y los atributos del elemento. En consecuencia, los datos de elección pueden ser una fuente precisa de la que se pueden aprender las preferencias. El modelo basado en la elección parece proporcionar una compensación óptima entre la interpretabilidad y el rendimiento y allana el camino para la aplicación de modelos de toma de decisiones más complejos en el campo de los sistemas de recomendación.
- El segundo punto está relacionado con el desarrollo de un modelo basado en características cognitivas y la creación de usuarios de modelos basados en la elección. Por un lado, utilizamos el modelo de conjunto para la predicción. La comprensión del proceso de toma de decisiones de los humanos es la clave para predecir el comportamiento humano y puede desempeñar un papel clave en un sistema de recomendación. La evidencia de campos como la economía del comportamiento, la neurociencia y el neuromarketing sugiere que los circuitos del Sistema I impulsados por factores emocionales son los que controlan la mayoría de nuestras elecciones diarias. el análisis muestra: (1)

los modelos Sistema-I-Emocionales podrían no ser los mejores predictores del comportamiento de elección, (2) no existen elementos como el mejor modelo de elección o el mejor conjunto para todos los usuarios, y (3) los conjuntos pueden funcionan mejor que los modelos de opción única, pero se debe encontrar un método de construcción óptimo.

– En tercer lugar, consultamos los métodos óptimos para construir conjuntos basados en elecciones. Exploramos las posibilidades de métodos ciegos o desinformados e informados. El primer método se enfoca en dos estrategias para construir la agregación de estudiantes: conjuntos de tipo 1-Learner y N-Learners, y el segundo método se concentra en crear un conjunto basado en información previa. Los resultados con los nuevos modelos son alentadores y motivan una mayor exploración de métodos elaborados para construir mejores conjuntos basados en la elección. El estudio también sugiere que un método informado es un método poderoso para lograr predicciones aún más precisas. Creemos que este es un camino emocionante para explorar el funcionamiento interior de los cerebros humanos.

A pesar del trabajo realizado en la tesis doctoral, aún queda mucho trabajo por hacer. Para mejorar y generalizar los resultados relacionados con los modelos propuestos, nuestro trabajo futuro se centrará en la replicación de los experimentos, incluida la selección de otras metodologías y diferentes métricas. La tesis se centra en el dominio del turismo y el dominio del cine / entretenimiento; para futuros estudios, planeamos aplicar el estudio en el dominio del mercado.

# Contents

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction.

Recommender Systems (RSs) are software tools and methods presenting suggestions for items to be of worth to a user [36]. To do so, RSs gather information on the preferences of the users for a set of items, which can be got explicitly or implicitly. They also can use users' demographic features [9]. In another meaning, the task of RSs is to convert data on users and their preferences into predictions of users' possible future likes and interests [30]. The proposals provided by an RSs are intended to support their users in various decision-making processes, besides being a valuable means to cope with information overload. Several techniques for recommendation generation have been suggested.

Development of RSs is a multi-disciplinary effort that includes specialists from different fields, such as artificial intelligence, data mining, statistics, decision support systems, marketing, human-computer interaction, consumer behavior, and psychology [36]. As a result, they can be used in different fields, and by collecting information available on the web, the user's recommendation becomes personalized. To make that, the process building an internal representation of a user based on the data collected about the users' characteristics such as age, gender, interests, preferences, etc. Several levels of personalization could exist within a system using explicit user characteristics (e.g., age, gender, demographics) or implicit user behavior pattern (web browsing history, click model). Many hidden variables, such as personalities, emotions, and moods, which typically are not explicitly provided by users can be deduced from user-generated content [41].

This new generation of “human” recommenders, also called personality-based recommender systems [42, 21, 15, 33], should consider the “character” of users as something more than simply individual behavior, but an extensive and complex group of ideas, feelings, and behaviors.

The RSs problem can be approached in different ways by viewing it as the problem of predicting user’s choices in any particular context. Under this viewpoint, the Rational Choice Theory can be considered the classic model used to explain the choices made by rational agents [40]. In order to predict users’ choices, we need to understand the decision-making process of humans. The state-of-art about DM is controlled by the Dual Process Theory that considers two main types of cognitive processes [46]: The first one is fast, unconscious and effortless; the second one is slow, conscious and energy-consuming. These processes were associated to cognitive systems by psychologist Daniel Kahneman who coined the terms System I and System II [25].

Hence, to construct recommenders even more successfully, we must combine technical designs for recommender applications with deep knowledge about human decision-making processes [23]. The following motivations came after a thought analysis of the state-of-the-art:

1. The fact that there is an increasing need to explain why a website is recommending some items and not others.

2. The limitations found in significant paradigms of current recommender algorithms in terms of the accuracy of the learned preferences.

3. The need to provide an interpretation of the generated predictions, in other words, how to explain the inner workings of the algorithms in a more transparent way.

4. The intuition that understanding and modeling of human decision-making could be an interesting approach to develop algorithms to predict what humans would choose if they would have access to all the relevant information.

## 1.2 Objectives

As described above, the main goal of this thesis is TO PRESENT NEW TECHNIQUES AND METHODS TO DEVELOP RECOMMENDER SYSTEM ALGORITHMS based on the use of the choice-based models and applied its ensembles models to provide more accuracy and explained the decision-maker behavior. In order to test our systems, two datasets used, first, the Gastronomic dataset, which focuses on the tourism domain, second, MOVISTAR dataset, WHICH FOCUSES ON THE MOVIE/ENTERTAINMENT DOMAIN.
The specific objectives of the thesis are as follows:

1. To explore the potential and analyze the performance of choice-based models. The concept is borrowed from choice and utility theory, in which the user choices are the key data used to learn about the decision-maker preferences.
2. To develop cognitive models on the basis of general choice-based models. By resorting to the Dual-Process Theory, we have come up with System I as well as System II cognitive models.
3. To propose the hybridization of high-performance ensembles with transparent and interpretable choice-based models. Specifically, we have wondered about the optimal methods to build choice-based ensembles. More specifically, we analized uninformed, or blind, as well as informed methods to select the set of learners in the ensemble in the most efficient way.

## 1.3 Questions And Hypotheses

In this Ph.D. thesis, we address the primary objective by decomposing it into three objectives that cover several aspects of choice-based ensembles for a recommendation. To achieve these objectives, a set of research questions and hypotheses that are related to the objectives are formulated.

*The first questions regard with the first specific objective. They come from two observations in the domain of Recommender Systems: (1) the rating-based models assume a strong relationship between preferences and ratings, (2) and people are increasingly reluctant to supply explicit recommendations as they perceive them as an outcome of an advertising model.*

*We used choice-based models to overcome these problems.*

**Q1:** Which data is more relevant to generate recommendations: ratings or choices? **H1** Users choices seem to be the key data used to learn about decision-makers preferences.
**Q2:** What is the performance of choice-based models compared to rating-based algorithms and basic ensembles? **H2** Choice-based algorithms are expected to overcome the performance of both rating-based algorithms and basic ensembles.

*The next set of questions regards with the second specific objective. We have developed three different types of cognitive models: emotional, attentional, and rational. The later can be classified as a pure System-II model, while the first and second ones could be cast as System-I models. We then posed some questions about their performance.*

**Q3**: What is the best cognitive choice-based model: S-I-emotional, S-I-attentional, or S-II-rational? **H3.**: On the basis of the evidences from the fields of behavioral economics and neuromarketing, we believe that S-I models perform better than S-II models. More concretely, considering the somatic marker hypothesis [13], we expect the S-I-emotional model to be the best candidate to predict human choices.
**Q4**: Would it make sense to talk about the best cognitive choice-based model for all users? **H4.**: Yes. By following the same argument of the previous point, we think that S-I-emotional models will be the best predictors for all users.
**Q5**: Could the performance of the best single choice-based model be overcome by using an ensemble of rational-emotional-attentional learners? **H5.**: Yes. According to the state-of-art in machine learning, ensembles usually provide the best performance by taking advantage of the aggregation of information from different learners [34, 27].
**Q6**: Would it be correct to talk about the best ensemble for all users? **H6**: Yes. It is expected that ensembles built on S-I-emotional features will be the best among all users.

*Afterward, we focused on the third specific objective, and we applied uninformed or blind methods to build choice-based ensembles. Therefore, it was assumed that no prior information about the performance of the learners of the ensemble is available. The specific questions and hypotheses that have guided this part of the work are briefly described:*

**Q7**: Which method is the best to build choice-based blind ensembles?: Two strategies have been studied: 1-Learner and N-Learners type ensembles. **H7**: Considering diversity as a key factor of an ensemble's performance, the N-learners scheme should provide the best results.
**Q8**: Will choice-based ensembles outperform single choice models? **H8**: We expect the best choice-based ensembles to perform better than the corresponding single models.
**Q9**: How good are choice-based ensembles compared to ensembles built with other models? As the baseline model, we have resorted to Decision Trees (DT), which provide transparency and interpretability. **H9.**: We predict that blind choice-based ensemble will show better performance than DT-based ensembles.

*Finally, we completed the study about the third objective by exploring informed methods to build the choice-based ensembles. Therefore, it was assumed that prior information about the performance of the learners of the ensemble is available. The specific questions and hypotheses that have guided this work are briefly described:*

**Q10**: Which method is the best to build informed choice-based ensembles? Three types of prior information have been studies Diversity, MSE, and Crowd Error. **H10**: Considering diversity as essential prior information of ensembles performance, the diversity should provide the best result.
**Q11**: Will informed choice-based ensembles method outperform a single choice model? **H11**: We expected the best informed choice-based ensembles to perform better than the corresponding single models.
**Q12**: How good are choice-based ensembles compared to ensembles built with other models? As the baseline model, we have resorted to Decision Trees (DT), which provide transparency and interpretability. **H12**: We predict that informed choice-based ensemble will show better performance than DT-based ensembles.
**Q13**: Which method the best to build optimal choice-based ensembles? Two methods have been studies: Blind on uninformed and informed ensembles. **H13**: We expected the informed ensemble would show better performance than the blind ensembles.

## 1.4 Contributions

This Ph.D. thesis makes various scientific contributions that permitted us to achieve the main goal and might facilitate an increase in the use of choice-based models in the recommender systems. These contributions are as follows:

- First, the use of a choice-based model approach may solve the point of the accuracy of the preferences learned from ratings. While ratings describe a post-experience outcome that may be more dependent on the satisfaction of previous users expectations rather than on preferences, choices are the result of the direct matching between the user's preferences and the item's attributes. Consequently, choice data may be an accurate source from which preferences might be learned. The choice-based model seems to provide an optimal trade-off between interpretability and performance and paves the way to the application of more complex decision-making models in the field of recommender systems.
- The second point is related to the development of a model based on cognitive features, and building choice-based model users. On the one hand, we used the ensemble model for the prediction. The understanding of the decision-making process of humans is the key to predict human behavior and can play a key role in a recommendation system. The evidence from fields like behavioral economics, neuroscience and neuromarketing suggest that System I circuits driven by emotional factors are the one in control of most of our daily choices. The analysis shows: (1) System-I-Emotional models might not be the best predictors of choice behavior, (2) there are no such things as the best choice model or the best ensemble for all users, and (3) ensembles may perform better than single choice models, but an optimal building method has to be found.
- Third, we query about the optimal methods to build choice-based ensembles. We explore the possibilities of blind or uninformed and informed methods. The first method is focusing on two strategies to build the aggregation of learners: 1-Learner and N-Learners type ensembles, and the second method is concentrating on raising ensemble based on prior information. The results with the new models are encouraging and motivate further exploration of elaborated methods to build better choice-based ensembles. The study also suggests that an informed method is a powerful method to achieve even

more accurate predictions. We believe this is an exciting pathway to explore the interior workings of human brains.

## 1.5 Thesis Structure

This Ph.D. thesis is organized into six chapters, including the current one.

- **Chapter 1**: Introduction which highlights the objectives of the study, motivation of the study, and hypotheses of the study, Contributions of the study, Publication list, and this section.
- **Chapter 2: State of the Art.** In state of the art, we argue of the different aspects related to this thesis are included. We discussed the previous studies related to recommender systems, choice-models, ensembles models, and decision-making models.
- **Chapter 3: Methodology.** In this chapter, we explained the experimentals design and the datasets that used, Ensemble Tree Methods, Choice models, Baselines Recommendation Methods, Cognitive Features For Choice Model, Optimal Ensemble Model For Choice, Optimal Ensemble Model For Choice, and Evaluation metrics.
- **Chapter 4: Results.** In this chapter, we have shown results on the experiments for (1) Ensemble Decision Tree And Baseline Recommendation Method, (2) Choice Model, Ensemble Decision Tree, And Baseline Recommendation Methods, (3) Choice Model Using Cognitive Features, (4) Optimal Choice-Based Ensemble Building By Emotion, Rational, And Attentional Features.
- **Chapter 5: Discussion.** The results were discussed and shown if it's confirmed and contradiction with the hypotheses.
- **Chapter 6: Conclusions.** In this section, we have shown our conclusions after analyzing the results and future works.
- **Bibliography.** It shows all the references that used in our research.

## 1.6 Publication List

In this section, all publications related to this thesis are listed. They have been categorized according to their type of conferences or journals. The publications that are submitted but not published are marked with (*).

**Conferences:**

- Paula Saavedra, Pablo Barreiro, Roi Durán, Ameed Almomani, Rosa Crujeiras, María Loureiro, Eduardo Sánchez Vila, Recommender Systems: machine learning vs. theoretical approaches, Machine Learning Workshop Galicia (CESGA, 2016).
- ALMOMANI, Ameed; SAAVEDRA, Paula; SÁNCHEZ, Eduardo. Ensembles of Decision Trees for Recommending Touristic Items. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, Cham, 2017. p. 510-519.27 May 2017. DOI: https://doi.org/10.1007/978-3-319-59773-752 [4].
- Almomani A., Monreal C., Sieira J., Graña J., Sánchez E. (2019) Rational, Emotional, and Attentional Choice Models for Recommender Systems. In: Rocha Á., Adeli H., Reis L., Costanzo S. (eds) New Knowledge in Information Systems and Technologies. WorldCIST'19 2019. Advances in Intelligent Systems and Computing, vol 931. Springer, Cham, DOI : https://doi.org/10.1007/978-3-030-16184-253 [3].
- ALMOMANI, Ameed; SÁNCHEZ, Eduardo. Uninformed Methods to Build Optimal Choice-Based Ensembles. In: International Work-Conference on the Interplay Between Natural and Artificial Computation. Springer, Cham, 2019. p. 58-65. DOI: https://doi.org/10.1007/978-3-030-19651-66 [5].

**Journals:**

- Almomani Ameed, Monreal Cristina, Sieira Jorge, Graña Juan, Sanchez Eduardo, "Rational, Emotional, and Attentional models for Recommender Systems", Expert Systems.DOI: 10.1111/exsy.12594.

– (*) Almomani Ameed, Saavedra Paula, Barreiro, Pablo, Durán Roi, Crujeiras Rosa, Statistics Loureiro María, Sanchez Eduardo, “Choice-based recommendation models: application in Tourism”, Expert Systems, submitted in 2019.

– (*) Ameed Almomani and Eduardo Sanchez,” Blind methods to build choice-based ensembles”, Natural Computing, submitted in 2019.

# CHAPTER 2

# STATE OF THE ART

The research carried out in this PhD is focused on three main paradigms: Recommender Systems, Decision Making, and Ensembles. In what follows, the main concepts and ideas of these paradigms are outlined.

## 2.1 Recommender Systems

### 2.1.1 What are Recommender Systems?

Recommender Systems (RSs) are software tools and techniques aimed at reducing information overload and providing suggestions for items that are most suitable for particular users [36, 11]. People are using these systems to find the information most relevant to them (Figure 2.1).

The recommendation process involves the gathering of user's information during system-user interaction, which is then used to deliver relevant content, services, and predict decision-making outcomes in order to satisfy the user's needs. The information can be collected explicitly by means of users' ratings, or implicitly by observing users' behavior, such as applications downloaded, movies watched, websites browsed, and books read. Furthermore, user's data may be enriched with demographic features of users, for instance, age, nationality, and/or gender. The final purpose is to improve the user's experience of a service.


**Figure 2.1:** Recommender Systems strategy.

Recommender systems support a broad range of applications, including suggesting appealing items (movies, books, music, etc) to buy, online news to read, and relevant search results. Popular companies like Google, Microsoft, Netflix, and Amazon, have developed their own systems for learning about preferences, tastes and human behaviours. Such systems are used on a daily basis to put offers on the spotlight, adapt product catalogues, and predict shopping baskets.

### 2.1.2 Popular examples

#### 2.1.2.1 Netflix

Netflix[1] is a media platform that provides personalized recommendations to help users to find relevant shows and movies. Whenever users access the Netflix service, a recommendation system aims to help users find a movie or discover the relevant ones with minimal user effort. Several factors are applied to estimate the likelihood that the user will watch a particular title in the catalog. For instance, (1) user's interactions with the service (such as user viewing history and the rating of experienced items), (2) experience of other users with similar tastes and preferences, and (3) information about the consumed items, such as their genre, categories, actors, release year, etc. In addition, the platform collects the time of the day that users watch,

[1] www.netflix.com

the devices used to watch Netflix, and the viewing time per session. All these data are used as inputs to feed the recommendation algorithm. The system does not incorporate demographic information (such as age or gender) as part of the decision making process.

To deal with cold start problems, Netflix uses “Jump starting”. When the user creates a Netflix account, or adds a new profile in the account, the user is asked to highlight some titles that he/she likes. These titles are used to “jump start” user’s recommendations. If they decide to skip this step, then the users will start with a diverse and popular set of titles. Once the user begins watching titles on the service, this will overwrite any initial preferences which were provided. As the user continues to watch over time, the titles watched more recently will outweigh titles watched in the past. In addition to decide which titles are going to fill the rows of a user’s Netflix homepage, the system also ranks each title within each row (Figure 2.2). The ranking offers a personalized catalog to each user. On each row, there are three tiers of personalization: (1) the choice of row (e.g., Continue Watching, Trending Now, Award-Winning Comedies, etc.), (2) which titles appear in the row, and (3) the ranking of those titles. To improve the Netflix recommendations system, they take feedback from every visit to the Netflix service. They continuously re-train the algorithms to increase the accuracy of their prediction of what users most likely to watch.

#### 2.1.2.2 Amazon

Amazon[2] focuses on recommending products that should be of interest to the customer while he/she is browsing the site. Amazon generates an information page for each item, presenting details of the item as well as purchase information. The "Buyers who Bought..." feature is included in the information page for each item on their list. Two separate recommendation lists are shown: (1) items that are commonly purchased by customers who bought the selected item, and (2) frequently purchased items of the same author as one of the chosen item [38].

Customers rate items on a five point scale, ranging from “hated it” to “loved it.”. The item *Matcher* feature enables customers to provide direct feedback about items they have read. Thereafter, Amazon uses the customer’s *Comment* feature and 1-5 star ratings to evaluate items as well as to obtain item recommendations based on the views of other customers. Found on the information page for each item is a list of 1-5 star ratings and written comments

[2] www.amazon.com

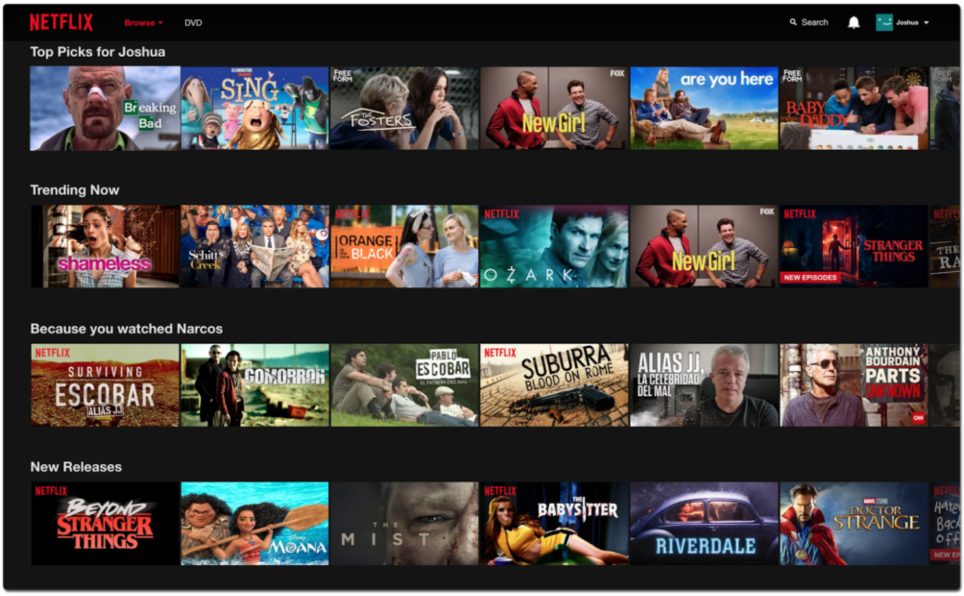


**Figure 2.2:** Netflix Web Page.

presented by customers who have read the item in question and submitted a review. Customers have the option of using these recommendations in their purchase decision.

## 2.2 Classification of Recommender Systems

Recommender systems are usually categorized according to the estimation procedure applied to predict an item's utility [6]. We list here the existing main methods:

- Content-based Filtering (CBF): The user will be recommended items similar to the ones the user preferred in the past.
- Collaborative Filtering (CF): The user will be recommended items or things that people with similar flavors and preferences preferred in the past.
- Hybrid Approaches: These methods merge collaborative and content-based approaches.

### 2.2.1 Content-based Filtering.

The field of Machine Learning has played an essential role by providing algorithms that work at the back end of the RSs. The first family of algorithms was referred to as content-based as they worked both with item attributes and with past user experiences to learn a profile of preferences for each decision-maker [6]. First, a user's profile is learned by analyzing the information of items previously consumed and rated. Second, the process follows by predicting the utility of all available items after matching the user's profile with the set of features of each item. For instance, in a movie recommendation utilization, to recommend movies to user X, the content-based filtering recommender system tries to understand the commonalities among the movies that user X has rated highly in the past (specific genres, actors, directors, subject, etc.). Next, only the movies that have a high degree of similarity with the user's preferences would be recommended. This method presents advantages, such as user independence as CBF systems only use ratings of the active user to create her model, but also some important limitations [6, 1].

- Limited Content Analysis. Content-based filtering is limited by the features that are explicitly associated with the objects that these systems recommend. Therefore, to have enough set of features, the content must either be in a pattern that can be analyzed automatically by a computer (e.g., text), or the elements should be assigned to items manually. Content-based filtering has difficulty in distinguishing between high-quality and low-quality information that is on the same topic. Furthermore, as the number of items grows, the number of items in the same content-based category increases, further decreasing the effectiveness of content-based filtering.

- Overspecialization. A second limitation, which has been studied widely both in this field and in others, is overspecialization. When the system can only recommend items scoring extremely against a user's profile, the user is restrained to seeing items similar to those already rated.

- New User Problem (Cold-start ). The user must rate a sufficient number of items before a content-based filtering RSs can truly understand the user's preferences and give the user with reliable recommendations. Hence, a new user, having very few ratings, would not be able to get reliable recommendations.

What is more, the need to adapt the algorithms to manage different business domains, with disparate types of products, have revealed the practical limitations of the content-based approach. This led to the development of a second family of algorithms: collaborative-based filtering.

### 2.2.2 Collaborative-based Filtering

Unlike content-based recommendation methods, collaborative-based filtering relies on user's ratings to generate recommendations. The main idea is to persuade users to rate consumed items, and use later the set of available ratings to predict the rating of any new user-item interaction [2, 17]. For instance, in a movie recommendation utilization, to recommend movies to user X, the collaborative-based filtering method tries to find its "social matches", i.e. other users that have similar tastes in movies (rate the same movies similarly). Then, only the movies that are most liked by the "matches" of the user X would be recommended. The main idea of collaborative-based filtering methods is that the unknown ratings can be estimated because the observed ratings are often highly correlated across different users and items. There are two types of methods that are commonly used in collaborative-based filtering: memory-based, and model-based.

#### 2.2.2.1 Memory-based Methods.

Memory-based methods are heuristic techniques that predict ratings based on the entire collection of previously rated items by the users. Also known as neighborhood-based collaborative filtering algorithms, these methods make the prediction by using the set of ratings of each user's neighborhood. Memory-based methods can be classified in one of two ways:

– User-Based algorithms.

  User-Based algorithms employ the whole user-item database to generate a prediction. The user's neighborhood is identified as the set of users that have a history of matching with the active user. More specifically, the neighborhood is made up with those users providing similar ratings to the ones of the target user. It is assumed that those users will share the same interest and will thus have similar preferences. Once a neighborhood is formed, these algorithms use different procedures to combine the preferences of neighbors to produce a prediction or top-N recommendation for the active user. The process of a typical user-based collaborative filtering would operate as follows [20]:

1. Similarity computation: The recommender system computes the similarities between the active user and other users who have rated the same items.
2. Neighborhood: Similar users are regrouped in a subset, namely the neighborhood of the active user. Typically, this neighborhood is made up with those users in the top list of similar users.
3. Prediction and recommendation generation: In the final phase, the system gathers information from the active user's neighborhood. Then an algorithm generates a list of items to be recommended, called the Top-N recommendation list, or a rating prediction for a specific item.

– Item-Based algorithms

  Item-based are alike to user-based collaborative filtering algorithms, but from an item point of view. Items that have been rated in the same way are likely to share some similar characteristics, thus users who like one of them should like the others that are similarly rated [19].

Collaborative based filtering systems have some problems [28, 6]. The first one is the New-User problem, the same issue found with content-based systems. To make accurate recommendations, the system must first learn the user's preferences from the ratings that the user gives. The second one, the New-Item problem, happens when new items are added regularly to the catalog. Collaborative systems rely only on users' preferences to make recommendations. Hence, until the new item is rated by a substantial number of users, the recommender system would not be able to recommend it. The third problem is the Sparsity. In any recommender system, the number of ratings obtained is usually relative small compared to the number of ratings that need to be predicted. Dimensionality reduction techniques, such as the Singular Value Decomposition (SVD), are used to reduce the dimensionality of sparse ratings matrices.

#### 2.2.2.2 Model-based Methods.

In model-based methods, data mining and machine learning methods are merged in the context of predictive models. In cases where the model is parameterized, the parameters of this model are learnt within the context of an optimization framework. Some examples of model-based methods include Bayesian Methods, Decision Trees, Rule-based Models, and Latent Factor Models. Multiple of these methods, such as latent factor models (ex. SVD, Matrix factorization), have a high level of coverage even for sparse rating matrices.

### 2.2.3 Hybrid Approaches.

Many recommendation systems use a hybrid approach by combining collaborative as well as content-based methods. Hybrid recommender systems have been designed to avoid specific limitations of content-based filtering and collaborative based filtering systems. i.e., they are not able to recommend items that have not been rated yet. However, the description of new items is usually available and can be used with content-based filtering methods. Several methods to develop hybrid recommender systems have been proposed:

1. Executing collaborative and content-based methods separately and combining their predictions.
2. Combining some content-based characteristics into a collaborative approach.
3. Incorporating some collaborative features into a content-based approach.
4. Building a general unifying model that combines both content-based and collaborative based features.

## 2.3 Decision Making Paradigms

We believe that the recommendation problem may be solved more accurately by facing it as a decision-making problem, and more specifically as a choice prediction task [37]. The current paradigm in the Recommender Systems field mainly focuses on rating prediction and item's recommendations based on the highest predicted value. However, this approach can easily assume unrealistic situations. For instance, recommendations could be based on predicted ratings on high-quality items that, considering budget limitations, could not reach the choice set of the users, and will thus never be chosen. In this PhD, it is argued that modeling actual choice behavior is a promising approach to overcome these limitations.

### 2.3.1 Discrete Choice models

Discrete choice models represent decision-makers' choices among alternatives. The decision-makers can be people, firms, or any other decision-making agent. The alternatives may belong to disparate domains, such as, competing products, courses of action, or any other options or items over which choices must be performed [43]. In order to suit within a discrete choice structure, the set of alternatives, called "the choice set," needs to satisfy three conditions. First, the alternatives must be mutually exclusive from the decision-maker's perspective, i.e the decision-maker chooses only one alternative from the choice set. Second, the choice set must be exhaustive, in that all possible alternatives are included. Third, the number of alternatives must be limited.

The standard tool for modeling individual choice behavior is the discrete choice model based on the random utility hypothesis. These models are based on the Behavioral Theory, which is a branch of Economy that is [7]: (i) descriptive, it assumes how human beings behave and does not prescribe how they must behave, (ii) abstract, it can be formalized in terms which are not specific to a particular circumstance, and (iii) operational, it generates models with parameters and variables that can be measured. Formally stated, a specific theory of choice is a collection of procedures involving the following elements [7]:

1. Decision maker. It can be a person or a group of persons. Individuals face different choices and have widely varying tastes. Therefore, the differences in decision-making processes among individuals must be explicitly treated.

2. Alternatives. A choice is made from a non-empty set of alternatives. Alternatives must be feasible to the decision maker and known during the decision process.

3. Attributes of Alternatives. The attractiveness of an alternative is evaluated in terms of a vector of preferences over the alternative's attributes. The preferences are measured on a scale of attractiveness.

4. Decision rule: The mechanisms used by the decision maker to process the information available and arrive at a unique choice. Random utility theory is one of a variety of decision rules that have been proposed and is the most used procedure.

Discrete choice models are ordinarily derived under an assumption of utility-maximizing behavior by the decision-maker. Traditional consumer theory assumes deterministic behavior, which states that the utility of alternatives is known with certainty and that the individual is always considered to select the alternative with the utility maximization. However, these assumptions have significant limitations for practical applications. Indeed, humans must handle the lack of information or ignorance about the available alternatives, which suggests that choice models should explicitly assume some level of uncertainty. The traditional consumer theory fails to do so. The concept of *random utility models (RUMs)* states the idea that modelling individual choice behavior is a probabilistic problem. The individual is still assumed to select the alternative that maximizes utility, but the analyst or modeller does not know the utilities with certainty. Therefore, the utility is decomposed in two parts: (1) an observed measurable element, and (2) an unobserved or unknown element that is modelled as a random variable. The models can also be seen as merely explaining the relation of explanatory variables to the result of choice, without indicating exactly how the choice is made.

*Logit* is by far the easiest and most widely used discrete choice model. Its popularity is because the formula for the choice probabilities takes a closed-form by assuming a certain distribution for the random variable and is readily explainable. It is derived under the assumption that ($\varepsilon$) is independent and identically distributed (i.i.d.) extreme value for all i. The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives. This assumption, while restrictive, provides a very suitable form for the choice probability. However, the assumption of independence can be unsuitable in some situations, and the development of other models has mainly arisen to avoid the independence assumption within logit.

*Mixed logit* is an extremely flexible model that allows the unobserved factors to follow any random utility model. It resolves the limitations of standard logit by allowing for random taste variations, free change patterns, and correlations in unobserved factors over time. It is not limited to normal distributions. Mixed logit derivation is straightforward, and simulation of its choice probabilities is computationally simple.

### 2.3.2 Cognitive Models.

The estimation of choice-based models require the identification of the relevant processes and variables underlying human decision-making (DM). The state-of-art about DM is dominated by the Dual Process Theory that considers two main types of cognitive processes [46]: the first one is fast, unconscious, and effortless; the second one is slow, conscious and energy-consuming. These processes were associated with cognitive systems by psychologist Daniel Kahneman who coined the terms *System I (S-I)* and *System II (S-II)* [25].

#### 2.3.2.1 System II and Rational Behavior

Traditionally, humans were considered rational agents and human choices were explained by means of modeling the rational processes of System II. S-II requires attention to the effortful mental activities that demand it, including complex computations. Processes such as the personal experience of agency, the conscious thinking about ourselves, the decision about what to think and what to do, all belong to the realm of S-II. This system is slow and can build thoughts in an arranged series of steps. The highly diversified operations of S-II have one feature in common: they need concentration and are interrupted when the brain is distracted. The following are examples of S-II:

1. Tell someone your phone number.
2. Park in a narrow space.
3. Fill out a form.

The fields of Rational Choice and Utility Theory joined forces to build a conceptual model of serious choices governed by S-II processes [39]. The mathematical description of humans being agents looking to maximize their own utility seemed to fit well with the behavior in the economic domain and also allowed to forecast the evolution of markets.

#### 2.3.2.2 System I and Irrational Behaviour

The essential idea of S-I is that it operates automatically and quickly, with little or no effort and no sense of voluntary control. S-I generates first impressions, and emotions are used as primary sources of sincere beliefs. The following are some examples of activities that are connected to S-I:

1. Detect that one object is more distant than another.
2. Answer to 3 + 3 = ?
3. Drive a car on an empty road.

All these examples are mental events that occur automatically and require little effort. Humans are born prepared to understand the world around us, recognize objects, orient attention. Other mental actions become fast and automatic by practice. S-I has learned associations between ideas (the capital of Spain?); it has also learned skills such as reading. Some skills are acquired only by specialized experts. Detecting the similarity of a personality perspective to a professional occupational stereotype requires a comprehensive knowledge of the language and the culture. The knowledge is cached in memory and accessed without intention and effort. Many of the mental activities on the list are entirely involuntary. You cannot avoid simple understanding sentences in your language or from orienting to a powerful, nor can you prevent yourself from knowing that 3 + 3 = 6 or from thinking of Madrid when the capital of Spain is mentioned. We cannot be turned off the S-I. For instance, If you are shown a word on the screen in a language you understand, you will read it unless you distracted your attention elsewhere. Sometimes, especially under uncertainty or in a complex context, these fast judgments are incorrect or flawed. This is why the behaviour derived from S-I evaluations is often termed as irrational.

The field of behavioral economics put S-I in the spotlight by revealing significant deviations between human behavior and the predictions of rational models [45]. A number of cognitive heuristics were identified suggesting that logical decision-making might not be as important as it was supposed. Herbert Simon described these findings as examples of ”bounded rationality”. In the last decade, System I models have gained major appeal in order to understand the processes behind a cognitive outcome[29]. In this context, fields like Neuromarketing or Neuroeconomy has grown in order to apply neuroscientific methods to gather relevant data [35]. EEG monitoring, for instance, constituted a useful source to record the neural activity of the cerebral cortex. The data is then used to map neural responses to emotional variables. Gaze monitoring and eye-tracking is another popular tool to track the degree of attention. The underlying rationale is that the scanpath is driven by attentional processes, and therefore a proxy for subject's preferences. By learning the features of the items that the user's is attending, we can then predict user's choice.

#### 2.3.2.3 Interplay between System I and System II

The key question about the Dual Process Theory is how both systems may work together efficiently. S-II, for instance, is engaged when you are asked to do something that does not come out naturally, and you will find that the regular support of a set demands a continuous effort. Everyone has some consciousness of the limited capacity of attention, and human behavior makes allowances for these limitations. When S-I encounters difficulty, it calls on S-II to carry more detailed and particular processing that may solve the problem at the moment. S-II is prepared when a question occurs for which S-I does not have an appropriate answer. The division of labor between S-I and S-II is extremely effective; it reduces effort and optimizes performance. One of the tasks of S-II is to overcome the desires of S-I. In other words, S-II is in charge of self-control. Because S-I runs automatically and cannot be turned off at desire, errors of intuitive thought are often difficult to prevent. Biases cannot always be avoided, because S-II may have no evidence of the when cues to likely errors are available. Even when cues that may signal dangerous stimuli are available, they can be prevented only by the improved monitoring and effortful activity of S-II. One of the main functions of S-II is to monitor and control decisions and behaviors suggested by S-I, allowing some to be expressed directly in behavior and crushing or changing others.

## 2.4 Ensembles and Ensemble-based Recommendations

Ensembles consist of aggregations of weak algorithms (aka learners) to produce more accurate predictions than the ones provided by a single learner [34, 18]. The purpose behind the ensemble methodology is to take advantage of several individual learners and combine them to obtain a learner that exceeds every one of them. In fact, human beings often consider several opinions, evaluate them, and aggregate the evaluations before making any significant decision [34].

Ensembles can be built by using either uniform or non-uniform schemes. Uniform ensembles are constituted from several instances of the same learning algorithm, which is known as a weak learner. Such learners are trained by using different samples or subsets of the training set. On the other hand, the non-uniform ensembles are built from running a set of different learning algorithms or learners over the complete training set. Such models ended up having different views or perspectives about the data [14, 44]. Ensemble models have become the dominant paradigm in the domain of Recommender Systems after the success of the Bellkor team in the Netflix prize [27].

### 2.4.1 Traditional Ensemble Methods

The most popular methods proposed for efficient aggregation of learners [34, 18] are: (1) bagging, (2) boosting, and (3) random forest.

- *Bagging*, acronym for bootstrap aggregating, is one of the first effective ensemble models. It is also one of the simplest to implement, with good performance, and can deal with both classification as well as regression problems. Bagging involves a set of weak learners, which are instances of a base learner, and are fitted in a parallel way. Diversity in bagging is achieved by using bootstrapped replicas of the training data, random samples of the training data that are randomly drawn with replacement from the entire training data. Each training data subset is used to train an instance of the same learner type. The individual learners are then aggregated by taking an average of estimation, or a majority vote of their predictions. Figure 2.3 presents an overall view of the bagging ensemble learning technique. For any given instance, the outcome chosen by the ensemble is the one that was majority chosen among the individual learners classifiers.
- *Random forests* are exceptional cases for bagging, constructed from decision trees by an aggregate of tree predictors. Each tree relies on the values of a random vector sampled separately. Such parameters can be bootstrapped replicas of the training data, as in bagging, but they can also be different feature subsets as in random subspace methods and with the same pattern for all trees in the forest.
- *Boosting* is a meta-algorithm that can be seen as a model averaging method. It is a widely used ensemble method, designed for classification and regression. Boosting also builds an ensemble of classifiers by resampling the data like the bagging. First, it creates a weak classifier, which satisfies that its accuracy on the training set is only slightly better than random guessing. A sequence of models is thereafter built iteratively, each

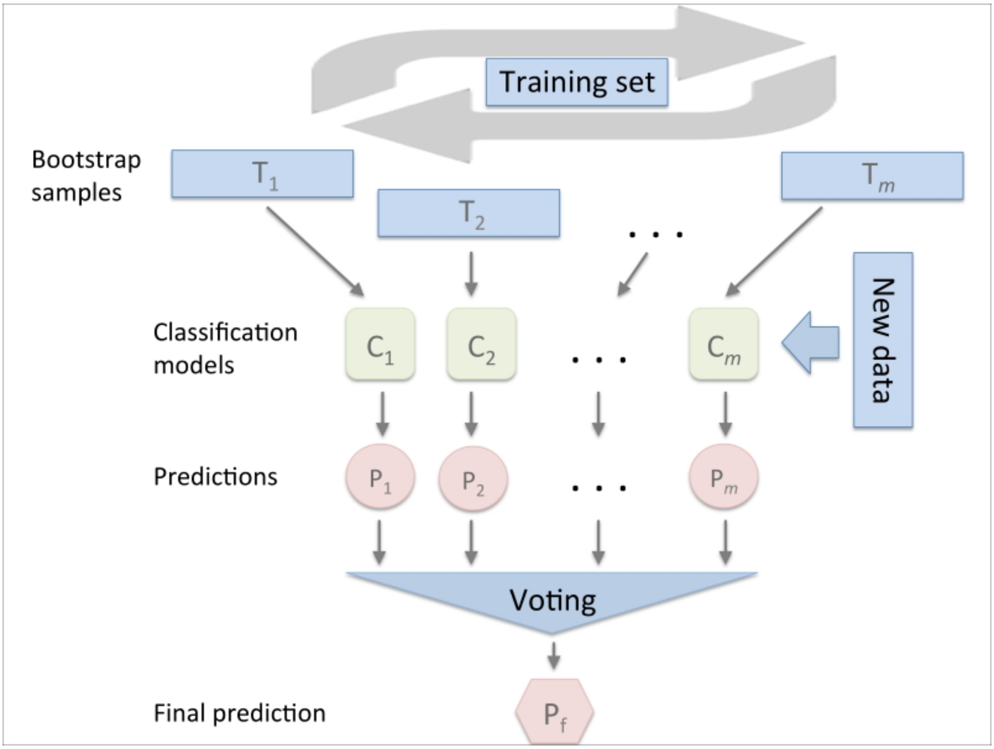


**Figure 2.3:** Bagging ensemble with majority vote as the aggregation technique.

one trained on a data set made from either misclassified observations or observations that were poorly predicted in the first step. After that, all of the sequential models are weighted according to their success. Finally, the outputs are aggregated using majority voting for classification, or average of estimation for regression.

Regard to the Non-Uniform ensemble models, different learners are used to constructing the ensemble, and all of them share the same dataset during the training stage. Different classifiers are then combined, the final prediction computed by an average of estimation or a majority vote of their predictions.

### 2.4.2 Reasons for Using Ensemble Methods

An ensemble approach may be used for various theoretical and practical reasons [34]:

- *Statistical Reasons:* Different classifiers with similar training performances may have multiple performances. Even classifiers with similar performances may work differently in the field, mainly if the test dataset used to determine the performance is not enough representative of the future field data. In these cases, aggregated the outputs of different classifiers by averaging may reduce the risk selection of a poorly performing classifier. The averaging may reduce the overall risk of making an especially poor selection.
- *Large Amounts of Data:* In particular applications, the volume of data to be analyzed can be too large to be efficiently manipulated by a single classifier.
- *Too Small Data:* Ensemble systems can also be used to deal with too little data problems. The amount of training data is essential for a classification algorithm to learn the underlying data distribution successfully. In the lack of sufficient training data, resampling techniques can be used for creating random subsets data, each of which can be used to train a different classifier, and building the ensemble models.
- *Divide and Conquer:* Despite the amount of available data, some problems are difficult for a classifier to solve. Mainly, the decision boundary that separates data from different classes may be too complicated or lie outside the scope of functions that can be performed by the individual classifier model.
- *Data Fusion:* If we have different sets of data collected from various sources, where the type of features are different, a single classifier cannot be used to learn the information held in all of the data.

### 2.4.3 Advantages and Disadvantages of Ensemble Methods

To summarize, ensembles have proven useful in many areas of machine learning as well as in the domain of Recommender Systems. In what follows, the main pros and cons are listed:

- Advantages of Ensemble Methods:

1. Often generate more accurate prediction results.
2. Provide stable and more robust models. The combined result of various models is always less noisy than the single models.
3. Allow handling different needs of difficult problems.
4. Provide additional degrees of freedom to manage the bias/variance tradeoff.
5. Solve the overfitting problem.

– Disadvantages of Ensemble Methods:

1. If the data come from a linear manner, linear models will be much better formed than ensemble models.
2. They suffer a lack of interpretability. So, the explanations of the predictions are seriously complicated, and make it difficult to explain why an algorithm delivers a certain recommendation.
3. Ensemble methods are computationally expensive. Hence, they require high learning times as well as large storage capacity.

# CHAPTER 3

# METHODOLOGY

## 3.1 Experiments.

### 3.1.1 Rectur

#### 3.1.1.1 Setup

We used the dataset gathered from an ecological experiment, named as the Santiago(é)Tapas experiment, carried out under the scope of the RECTUR project. The chosen setting was the fourth edition of the Santiago(é)Tapas contest, a gastronomic event that takes place every year in the city of Santiago de Compostela. For the event, 56 local restaurants proposed and elaborated up to three tapas that were sold at a fixed price. A total of 5517 participants, including local, Spanish and international users, tasted the available tapas over a period of 2 weeks. A TapasPassport was made available to all participants and included the following official information: (i) the contest guidelines, (ii) restaurant location, and (iii) the tapas offered at each restaurant. After consuming the tapas, participants evaluated their experience by providing a vote with two ratings (Figure 3.1): (i) a rating for the tapas, and (ii) a rating of the overall experience (service, place atmosphere, etc.).

It is important to point out that the experiment was carried out in a real setting rather than a laboratory setting. Thus, the restaurants were free to offer whatever type of tapas they wished, and the participants made their own decisions about which tapas to try.


**Figure 3.1:** RECTUR experiment. Images of votes, participating locals and TapasPassport.

#### 3.1.1.2 Design

The experiment gathered tapa's choices and ratings in order to create a recommendation system in the local gastronomy for tourists. As a tool to support the data management, a website[1] was created. Figure 3.2 shows the registration page used to get specific information about the participants of the experiment. Registration was promoted in the participating restaurants and in the Tapasporte during the Santiago(é)Tapas contest. Those who registered to participate in the experiment were gifted with 5 free tapa tickets.

The first 75 registered users were selected for the experiment. These people were then informed about the steps of the experiment. They were expected to try seven tapas among the ones participating in the contest. After consuming the tapas, their experience feedback was entered into a database. The database was also populated with the experience of other international and domestic tourists who tried and evaluated tapas participating in the contest. After gathering the data, the selected users were given five recommendations for tapas and restaurants generated by five different recommender algorithms.

[1] http://gsi.dec.usc.es/santiagoetapas/

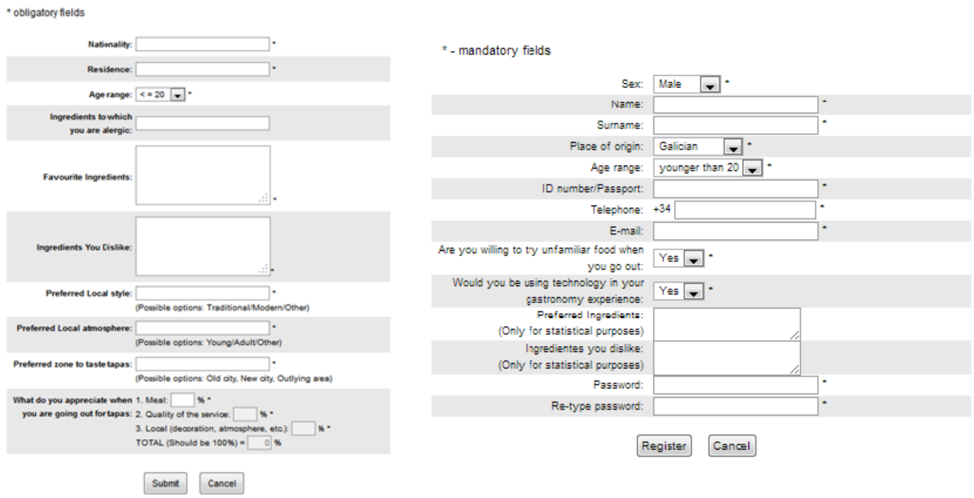


**Figure 3.2:** The registration pages for the Santiago(é)Tapas experiment in 2010 (left) and 2011 (right). The registration pages captured data about the participants of the Santiago(é)Tapas content and our experiment.

Apart from the website-based survey, a paper questionnaire was designed for international tourists. The questionnaire included queries about preferences on restaurant attributes. The purpose of the questionnaire was to increase the sample of interrogated international tourists to get more insight into preferences on restaurant attributes.

### 3.1.2 Movistar

#### 3.1.2.1 Setup

The experiments were carried out at CITIUS (Research Center on Information Technologies at USC) in collaboration with two Business partners: Neurologyca and Movistar. The first one is a Neuromarketing company that provided the experimental setup used to record subject's activity. The second one is the largest Telco in Spain that contributed with the stimuli presented to the subjects.

The experimental setup was aimed at monitoring and recording neural as well as gaze activity in order to find the drivers of human decision-making. Figure 3.3 shows the setup that consisted on a set of recording devices plus a manager application that facilitates the control


**Figure 3.3:** The experimental setup with the set of recording devices.

of the experimental procedure. The devices, described in Table 3.1, were placed in a room in which both temperature and illumination were controlled. With this setting, two main types of activity was recorded:

– Emotional. Electroencephalography (EEG) and Facial Coding (FACET) devices provided emotional variables. EEG records the electrical activity generated at the cerebral

**Table 3.1:** Devices and recorded variables.

| **Devices** | **Models** | **Variables** | **Variable Type** |
|---|---|---|---|
| EEG | Emotiv EPOC Headset | Frustration, Excitement and Engagement | Emotional |
| Facial Coding (FACET) | Logitech HD Pro webcam C920 | Joy, Anger, Fear, Surprise, Contempt, Disgust, Sadness, Neutral, Positive and Negative | Emotional |
| Eye-Tracking | TOBii X2 @30Hz | TimeSpend and Fixations | Attentional |
| iMotions (Software) | Version 6.2 | | |

cortex and the frequency bands are then processed to obtain the emotional variables. FACET works by identifying facial expressions and matching them with emotional or mood states.

– Attentional. The eye-tracker focused on recording the gaze activity. The parameters obtained, TimeSpend and Fixations, indicate the degree of interest or attention of each subject.

Further details about the recording of neural activity, face expressions, and eye movements are provided in the Appendixes.

#### 3.1.2.2 Design

The experimental task was a choice experiment in which subjects were presented a Web interface and asked to choose a movie from a set with four alternatives. The movies were labeled with three attributes (see Table 3.2) that were chosen by a relevancy criteria: Genre, Novelty and Price. The Novelty attribute indicated whether a movie is new (Release) or not (Catalog). As Novelty and Price were fully dependent, a Release Movie had always an associated cost, eight possible movie profiles could be created. Each profile was presented 10 times by each subject, so a total number of 80 movies were included on each trial. This set was finally structured in 20 choice situations with 4 stimuli on each one. So, a complete trial would generate 20 observations per subject (one for each choice situation).

The task protocol comprised the following steps: (1) Welcome, where the subject entered the room and asked to sign an informed consent; (2) Device Placement, where the recording equipment was placed on each subject and the corresponding calibration was carried out; (3) Directions, where the task was explained, and (4) Choice experiment, in which 20 choice situations with 4 movies on each set were presented trough the Web Interface (Figure 3.4).

**Table 3.2:** Characterization of movies: attributes and values.

| Feature | Values |
|---|---|
| Genre | Action, Comedy, Science Fiction, Drama |
| Novelty | Release, Catalog |
| Price | 4,99 euros (Release), 0 euros (Catalog) |

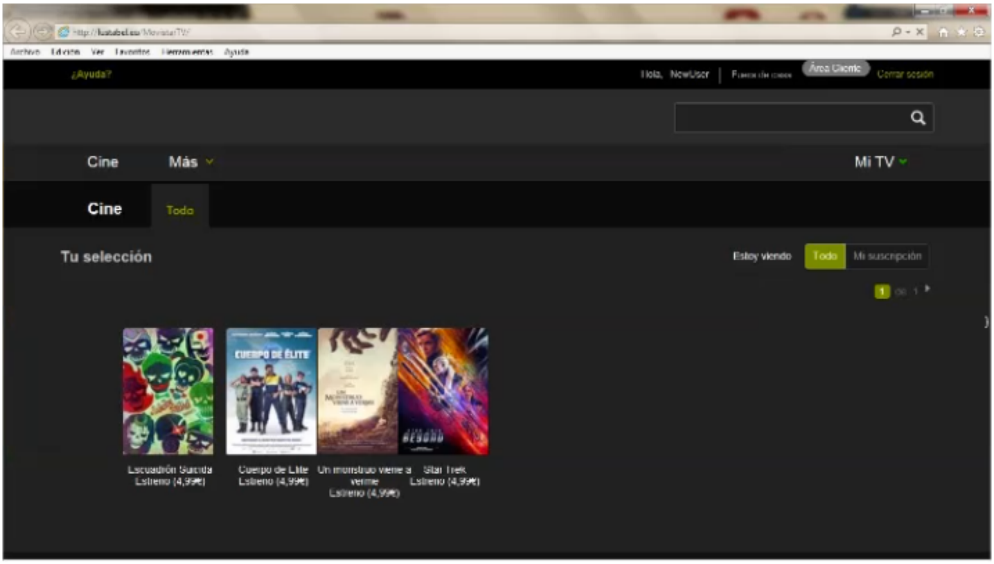


**Figure 3.4:** Web Interface.

**Table 3.3:** Tapa attributes and their corresponding values.

| **Attribute** | **Values** |
|---|---|
| Type | Cheese, Egg, Fish, Meat, Vegetable, Shellfish or Other |
| Character | Traditional or Novel |
| Rating | 0-5 |

The subject was asked to choose one of the four movies of the set and continue with the next choice problem.

## 3.2 Datasets

### 3.2.1 Rectur

The data collected in the experiment are summarized in the RECTUR data set. For each choice, the following tapa's attributes were recorded (Table 3.3): type, character and rating. Traditional tapas were created following well-known, popular recipes, while novel tapas were new and creative. Table 3.4 provides some relevant figures about the experiment.

**Table 3.4:** Experiment Information

| | |
|---|---|
| Participating restaurants | 56 |
| Different tapas offered | 109 |
| Tapas consumed | 35,000 |

### 3.2.2 Movistar

The dataset was collected during a choice experiment designed in collaboration with two Business partners: Neurologyca and Movistar. The subjects were presented a Web interface (Figure 3.4) and asked to choose a movie from a set of four alternatives across 20 trials or choice situations. The movies were labeled with three attributes (see Table 3.2): Genre, Novelty and Price. During the experiment both neural and gaze activity were recorded in order to find the explanatory variables of human decision-making. Two main types of variables were recorded as described in table 3.1.

Data were collected from 39 subjects (20 females, 19 males, age range: 18-51 years, nationality: Spain and Latin-American countries, selection sources: academia and business sectors). All subjects had prior experience using Internet and Web applications and had normal or corrected-to-normal vision. Due to sampling errors of EEG and FACET devices, some periods of the emotional activity were not properly recorded. This caused the number of valid observations, which may reach up to 20 (one for each choice situation), may be different for each subject.

## 3.3 Models

### 3.3.1 Choice-based models

The standard logit model and the mixed logit model, assuming Gaussian distribution on the coefficients, were chosen as basic representatives of the family of random utility choice-based models. Application of the mixed logit model was justified as we have found evidence of taste variations among decision-makers on the basis of both personal and contextual factors [22]. Due to the mutually exclusive requirement of the choice set, three different sets corresponding to the possible restaurant locations (old, new and outlying areas of the city) were established. Standard and mixed logit models were therefore estimated separately from these three choice subsets that include only the tapas associated with each zone.

In terms of data preparation, type and character attributes were transformed into eight dichotomous or binary variables associated with each value. To clarify the coding, we assumed that the decision-maker was located in the old town and chose tapa t100, made of meat type and novel character. The following considerations were taken to describe the choice: (1) the set of choices is the set of all tapas offered in the old town, (2) the "meat" variable will be set at 1 while the other type variables will be set at 0, and (3) the "novel" variable is set to 1 while "traditional" is set to 0.

#### 3.3.1.1 Recommendation as a choice problem

The recommendation problem can be approached in different ways by viewing it as the problem of predicting user's choices in any particular context. Under this perspective, the Rational Choice Theory can be considered the classic paradigm used to explain the choices made by rational agents [40]. This theory assumes that any decision-maker will solve the decision-making problem by applying the following rule:

$$CR(A, \succeq) = \{a' \in A \,\|\, a' \succeq a, \forall a \in A\} \tag{3.1}$$

where *CR* represents "choice rule" and the $\succeq$ operator represents the relationship "preferred to", or at least "preferred". The chosen alternative will therefore be that for which the decision-maker shows the greatest preference.

In order to build a predictive model on the basis of this rule, the researcher must replace the qualitative preference operator with a quantitative one that will enable numerical comparison between the benefit of each alternative. Utility theory comes to the rescue to solve this issue. One of its axioms states that it is possible to define a utility function such that,

$$a \succeq b \iff U(a) \geq U(b). \tag{3.2}$$

Therefore the choice rule in equation 3.1 becomes:

$$CR(A, \geq) = \{a' \in A \,\|\, U(a') \geq U(a), \forall a \in A\}. \tag{3.3}$$

This rule is mathematically equivalent to the formulation of the general recommendation problem [1], which is described in terms of a maximization problem:

$$a' = \arg\max_{a \in A} U(c,a) \Longleftrightarrow$$
$$CR(A, \geq) = \{a' \in A \,\|\, U(a') \geq U(a), \forall a \in A\}. \quad (3.4)$$

As the recommendation problem can be understood as a choice prediction problem, the powerful models and techniques developed in the latter field can be applied to generate recommendations.

#### 3.3.1.2 Discrete choice models with random utility

The choice rule represents how decision-makers reach their decisions. However, in the real world, researchers do not have access to all of the information that decision-makers may handle to estimate the utilities. For a specific user $c_n$, the researcher only knows some attributes of the alternatives, labelled $x_j$, for all $a_j$ alternatives with $j \in \{1, \cdots, J\}$. Therefore, the predicted utility can be decomposed as follows:

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad (3.5)$$

where $V_{nj} = V(x_j)$ is the representative utility, which can be estimated on the basis of the observed factors, and $\varepsilon_{nj}$ captures the unknown factors that cannot be observed by the researcher. This decomposition is fully general, as $\varepsilon_{nj}$ is defined simply as the difference between the true utility $U_{nj}$ and the representative utility $V_{nj}$.

The uncertainty about $\varepsilon_{nj}$ is handled as a random variable, and the researcher must make further assumptions about its probability distribution. The models derived under these assumptions are called random utility models (RUM) [31].

From the researcher's perspective, the choice rule of equation 3.3 for a decision-maker $c_n$, which is deterministic from the decision-maker's perspective, becomes probabilistic in the following way:

$$CR(A, \geq) = \{a_i \in A \,\|\, \mathbb{P}_{ni} \geq \mathbb{P}_{nj}, \forall a_j \in A\} \quad (3.6)$$

and the probability $\mathbb{P}_{ni}$ is estimated by considering the decomposition formulated in equation 3.5:

$$\mathbb{P}_{ni}(U_{ni} > U_{nj} \text{ for all } j \neq i) = \\ \mathbb{P}_{ni}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \text{ for all } j \neq i). \quad (3.7)$$

If the joint density of $\varepsilon_n = (\varepsilon_{n1}, ..., \varepsilon_{nJ})$ is denoted by $f$, the cumulative probability can be rewritten as follows:

$$\mathbb{P}_{ni} = \int_{\varepsilon} \mathbb{I}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \text{ for all } j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (3.8)$$

where $\mathbb{I}$ is the indicator function, equalling 1 when the term in parentheses is true and 0 otherwise.

#### 3.3.1.3 Standard and mixed Logit models

Different models are derived depending on the density chosen, i.e. depending on the evidence or assumptions about the distribution of the unobserved portion of utility. The simplest and most widely adopted choice model is the standard logit model [31], which is obtained under the assumption that each unobserved portion of utility $\varepsilon_{nj}$ is distributed independently an identically. In this case, $f$ denotes the density for Gumbel distribution and the integral 3.8 takes a closed form with the following solution:

$$\mathbb{P}_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}. \quad (3.9)$$

This model estimates the probability on the basis of the representative utility. If $V_{ni}$ increases, reflecting a match between the observed attributes of the alternative and the preferences of the decision-maker, with $V_{nj}$ for all $j \neq i$ held constant, $\mathbb{P}_{ni}$ approaches one, and $\mathbb{P}_{ni}$ approaches zero when $V_{ni}$ decreases, as the exponential in the numerator approaches zero as $V_{ni}$ approaches $-\infty$.

The representative utility is usually specified as linear in the set of alternative attributes: $V_{nj} = \beta_{nj} \cdot x_j$, where $x_j$ is a vector including, as before, the observed variables of the alternative $a_j$, and $\beta_{nj}$ denotes the model coefficients vector describing the preferences of decision-maker $c_n$ for the attributes of the alternatives $a_j$. The preferences $\beta_{nj}$ (model coefficients) are estimated by fitting equation 3.9 to a data set of choices. The choice set must verify three properties. It must be finite, exhaustive (the decision-maker always chooses one of the alternatives) and mutually exclusive (the choice of one alternative necessarily implies not choosing any of the other ones).

The standard logit model cannot represent differences in tastes that are not related to observed characteristics [43]. Therefore, if taste variation is modelled as partly random, a logit model with random parameters should be considered instead. Thus, $\beta$ is now a vector of random coefficients that vary across decision-makers in the population with density $g$. This density is a function of parameters $\theta$ that represent, in the Gaussian case, the mean and covariance of the random coefficient in the population. The choice probabilities can be written as follows:

$$\mathbb{P}_{ni} = \int \left( \frac{e^{V_{ni}(\beta)}}{\sum_j e^{V_{nj}(\beta)}} \right) g(\beta|\theta) d\beta. \tag{3.10}$$

As the previous integral does not adopt a closed form, it must be evaluated numerically. Once the researcher specifies a distribution $g$ for the coefficients, the parameters $\theta$ maximizing the simulated log-likelihood must be estimated through simulation. The $R$ draws of the coefficients are then taken from $g$ and the logit probabilities are computed for each draw. The unconditional probability in equation 3.10, which is the expected value of the conditional probabilities, is estimated as the average of the $R$ probabilities determined previously.

### 3.3.2 Cognitive models

The choice-based models presented so far require the vector of features $x_j$ to be identified. In what follows, these features are described and the cognitive choice-based models derived from them are introduced.

#### 3.3.2.1 Observed features

The data gathered in the experiment were aggregated in the Neurologyca-Movistar Dataset. Each choice situation was characterized with the choice set (4 movies), the chosen movie, and a set of features describing the process. These features included both the attributes of the movies and the subject's variables recorded during the experiment. Three different classes

**Table 3.5:** Observed features.

| Feature class | Feature Types |
|---|---|
| Rational | Action, Comedy, Science Fiction, Drama, Release and Catalog |
| Emotional | Frustration, Excitement, Engagement, Joy, Anger, Fear, Surprise, Contempt, Disgust, Sadness, Neutral, Positive and Negative |
| Attentional | TimeSpend and Fixations |

of features were considered (Table 3.5): rational, emotional and attentional. These sets were used to fit System I and System II cognitive models.

#### 3.3.2.2 System I and System II cognitive models

According to the available features, three types of cognitive choice models were built: S-II-rational, S-I-emotional, and S-I-attentional. Table 3.6 lists all the models that were fitted and evaluated combining the features of the Neurologyca-MoviStar Dataset.

To preserve individual preferences, we fitted the set of all models for every single subject. The minimum number of valid observations to learn a model was set to 12, which reduced the number of eligible subjects available for the experiment to 18. Two problems were considered: computing the best model for each user among rational, emotional, and attentional models, and identifying the best average model for all users and ranking this model for every user.

#### 3.3.2.3 Illustrative example

In order to illustrate the notion of a single choice-based model, a simple example of the rational one is provided. First, subject V18 was randomly chosen and her personal dataset was used to fit the Rational model (see Table 3.5). The $\beta_{nj}$ coefficients corresponding to the rational feature class were estimated. Once the model is learnt, the probability of choice $\mathbb{P}_{ni}$ for the four alternatives of a choice situation, i.e the movies are shown on the screen, are computed. Table 3.7 shows the estimated values of the coefficients as well as the choice probabilities. Finally, the alternative with maximum probability, $a_3$ in this case, is the one predicted to be chosen.

**Table 3.6:** Cognitive models used in this work: S-II-Rational (R), S-I-Attentional (A), and S-I-Emotional (E).

| Model | Features |
|---|---|
| R | Action, Comedy, Fiction, Drama, Release, Catalog |
| A1 | Action, Comedy, Fiction, Drama, Release, Catalog, TimeSpend |
| A2 | Action, Comedy, Fiction, Drama, Release, Catalog, FixationTime |
| E1 | Action, Comedy, Fiction, Drama, Release, Catalog, Anger, Frustration, Negative, Fear |
| E2 | Action, Comedy, Fiction, Drama, Release, Catalog , Excitement, Joy, Engagement |
| E3 | Action, Comedy, Fiction, Drama, Release, Catalog, Frustration, Excitement |
| E4 | Action, Comedy, Fiction, Drama, Release, Catalog, Frustration Frustration, Excitement, Engagement |
| E5 | Action, Comedy, Fiction, Drama, Release, Catalog , Anger, Fear |
| E6 | Action, Comedy, Fiction, Drama, Release, Catalog, Fear, Contempt, Disgust, Sadness |
| E7 | Action, Comedy, Fiction, Drama, Release, Catalog, Joy, Anger, Fear, Surprise |
| E8 | Action, Comedy, Fiction, Drama, Release, Catalog , Frustration, Excitement, Engagement, Joy, Anger |
| E9 | Action, Comedy, Fiction, Drama, Release, Catalog , Joy, Anger, Fear, Surprise, Contempt, Disgust, Sadness |
| E10 | Action, Comedy, Fiction, Drama, Release, Catalog, Sadness, Neutral, Positive, Negative |
| E11 | Action, Comedy, Fiction, Drama, Release, Catalog , Frustration, Excitement, Engagement, Neutral, Positive, Negative |
| E12 | Action, Comedy, Fiction, Drama, Release, Catalog , Anger, Fear, Surprise, Contempt, Disgust, Sadness, Neutral, Joy, Negative |
| E13 | Action, Comedy, Fiction, Drama, Release, Catalog, Frustration, Excitement, Engagement, Joy, Anger, Fear, Surprise, Contempt, Disgust, Sadness, Neutral, Negative |

**Table 3.7:** Prediction with a single rational models: $\beta_{nj}$ coefficients and choice probabilities $\mathbb{P}_{ni}$ for the alternatives in the choice set. Two of the features (Drama and Catalog) are represented as intercepts.

| **Coefficient** | **Value** | **Choice Probability** | **Value** |
|---|---|---|---|
| B:(intercept) | 0.22 | P($a_1$) | 0.08 |
| C:(intercept) | 1.31 | P($a_2$) | 0.27 |
| D:(intercept) | 0.90 | P($a_3$) | 0.56 |
| Action | 0.61 | P($a_4$) | 0.09 |
| Comedy | -0.44 | | |
| Science Fiction | -0.81 | | |
| Release | -1.40 | | |

### 3.3.3 Choice-based Ensembles

In this section, two approaches to build ensembles of single choice models were used: (1) uninformed or blind methods, and (2) informed methods. The following sections will discuss these approaches.

#### 3.3.3.1 Uninformed methods

Two strategies were applied to build uninformed ensemble models: 1-Learner and N-learners. The specific details of both arrangements are described below.

– ***1-Learner Ensembles.***

  Under this approach, only one type of weak learner is chosen. Thereafter, a number of instances of the same learner are trained with different samples of the dataset. Diversity in this case is achieved by exposing the learners to different parts of the dataset. In our study, two data sampling strategies were considered:

  – ***Boosting*** builds the ensemble by using an iterative procedure over the same learner. This means that a single learner improves its accuracy on the basis of the performance of previous versions of himself.

  – ***Bagging*** estimates different versions of the same learner on $M$ different bootstrapped training data set.

– ***N-Learners Ensembles***

In this approach, not one but N different weak learner types are considered. Each learner is trained using the full training set, so all learners sample the same set of data (Figure 3.5). Diversity is thus introduced in the ensemble by the different nature of the learner set. Two possible ways to build N-Learners ensembles were considered: (1) Common Strategy, and (2) Personalized Strategy.

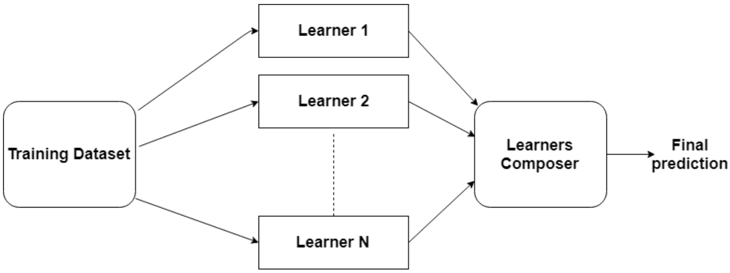


**Figure 3.5:** The Ensemble Model Framework.

1. ***Common Strategy***. It consists of identifying the best average Top-N learners for all users, and aggregating them to build a common ensemble. The procedure consists of the following steps:
   - Initial Learner Set. An initial set of all choice-based models is constructed. It contains the single choice-based models shown in Table 3.6: 1 Rational, 2 Attentional, and 13 Emotional models.
   - Model evaluation. All models are evaluated for each single subject, and then the average performance of every model for all subjects is estimated.
   - Ranking of models. A ranking based on the average performance is built.
   - Top-N model selection and common ensemble. The Top-5 models of the ranking were chosen to build a common ensemble for all users.
2. ***Personalized Strategy***. It consists on identifying the best Top-N learners for every user, and using them to build a personalized ensemble for each user. The procedure is described as follows:
   - Initial Learner Set. An initial set of all choice-based models is constructed. It contains the single choice-based models shown in Table 3.6: 1 Rational, 2 Attentional, and 13 Emotional models.

- Model evaluation. All models are evaluated for each single subject.
- Ranking of models. A ranking based on the average performance is built.
- Top-N model selection and personalized ensemble. The Top-5 models of each ranking were chosen to build a personalized ensemble for each user.

#### 3.3.3.2 Informed methods

Here we are focused on an informed method to build choice-based ensembles. Under this approach, some prior information about the performance of the individual learners is required before the ensemble is built. The metrics to measure such performance were: MSE, Crowd Error, and Diversity.

A **Greedy Ensemble selection** was used to choose each single learner to be included in the ensemble. The algorithm tries to estimate the globally best subset of learners (Table 3.6) by taking local greedy decisions for modifying the current subset [44]. Figure 3.6 shows an example of the search space for an ensemble of four learners using the greedy ensembles selection.

A different experiments were performed using the prior information, Low MSE, Low Crowd Error, or High Diversity :

- In the first experiment, when considering the diversity as prior information to build an ensemble, the first chosen model will be the best model, in terms of accuracy, either for each user (personalized) or all users (common). Afterward, we estimate the diversity that may be achieved by combining each of the possible second models with the first one. Then, the second model is chosen by picking up the one providing the highest diversity, and the ensemble is updated by adding the second and the first model. And we repeat the same steps to building the next ensemble model between best and picked a model with the next model, which has high diversity, these procedures repeated until building ensemble model for all models.
- In the second experiment, when regarding the MSE as prior information to build an ensemble, the first chosen model will be the best model, in terms of accuracy, either for each user (personalized) or all users (common). Then, we estimate the MSE that may be achieved by combining each of the possible second models with the first one.

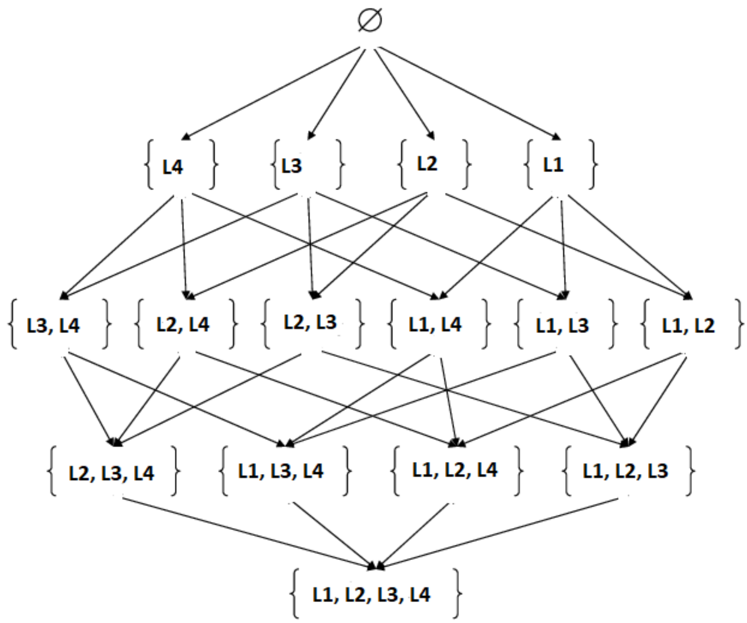


**Figure 3.6:** An example of the search space of Greedy Ensemble Selection algorithm for an ensemble of four learners.

Then, the second model is chosen by picking up the one providing the lowest MSE, and the ensemble is updated by adding the second and the first model. And we repeat the same steps to building the next ensemble model between best and picked a model with the next model, which has low MSE, these procedures repeated until building ensemble model for all models.

– In the second experiment, when regarding the Crowd Error as prior information to build an ensemble, the first chosen model will be the best model, in terms of accuracy, either for each user (personalized) or all users (common). Then, we estimate the Crowd Error that may be achieved by combining each of the possible second models with the first one. Then, the second model is chosen by picking up the one providing the lowest Crowd Error, and the ensemble is updated by adding the second and the first model. And we repeat the same steps to building the next ensemble model between best and picked a model with the next model, which has low Crowd Error, these procedures repeated until building ensemble model for all models.

Two possible approaches to build informed ensembles were studied: (1) Common Strategy, and (2) Personalized Strategy. We followed the same steps to identify the best model for all users (Common) and the best model for each user (Personalized), which mention in section 3.3.3.1, but we pick up the best ***one model*** in both strategies. then, we used the Greedy Ensemble Selection.

#### 3.3.3.3 Aggregation techniques

Under the bagging approach, two aggregation techniques were tested to estimate the final outcome: Probability Average, and Majority Vote. After testing that the results were similar with both techniques, only the Probability Average technique was implemented under the boosting approach.

– ***Probability Average*** For a decision-maker $c_n$, each learner of the ensemble estimates the choice probability for each item and then the average of probabilities is computed. The final prediction corresponds with the item with a higher probability average.

$$a' = \arg\max_{a_i \in A} \overline{\mathbb{P}_{ni}} \tag{3.11}$$

– ***Majority Vote*** The prediction using this technique is the most often predicted outcome by the set of K $C_k$ learners:

$$a' = \arg\max_{a_i \in A} \sum_{k:C_k(x_j)=a_i} 1 \tag{3.12}$$

### 3.3.4 Baseline Models

#### 3.3.4.1 Basic rating-based collaborative filtering

Two basic rating-based CF models were used: user-based collaborative filtering (CF-UB) and matrix factorization (CF-MF). CF-UB assumes that individuals with similar preferences will rate items in a similar way. Thus, missing ratings for a specific user $c_n$ can be predicted by finding a neighbourhood $N(n)$ of similar users and aggregating their ratings to calculate the corresponding prediction. The concept of similarity between users is used to define the neighborhood given all users within a similarity threshold. In this study, the cosine similarity measure was considered, and $|N(n)|$ was set at 25. For an item $i$ and an individual $c_n$, the ratings predicted, $\hat{r}_{ni}$, can be expressed as follows:

$$\hat{r}_{ni} = \frac{1}{|N(n)|} \sum_{j \in N(n)} r_{ji} \tag{3.13}$$

where $|\ |$ denotes the cardinal of $N(n)$.

CF-MF, on the other hand, characterizes both items and users by vectors of factors inferred from item rating patterns. For a given item $i$ and a user $c_n$, the vector $q_i$ measures the extent to which the item possesses those factors and the vector $p_n$, the extent of interest the user has in items that score highly on the corresponding factors. The dot product $q_i^T p_n$ captures the user's interest in the item's characteristics. This approximates user $c_n$'s rating of item $i$, $r_{ni}$, leading to the following estimate:

$$\hat{r}_{ni} = q_i^T p_n \tag{3.14}$$

Therefore, the challenge is to compute the mapping of each item and user to vectors $q_i$ and $p_n$. Here, singular value decomposition will be applied to factoring the user-item rating matrix, which may be sparse. In order to learn the factor vectors ($p_n$ and $q_i$), the regularized squared error on the set of known ratings is minimized:

$$\min_{q*,p*} \sum_{(u,i) \in K} (r_{ni} - q_i^T p_n)^2 + \lambda (\|q_i\|^2 + \|p_n\|^2) \tag{3.15}$$

where $K$ is the set of the $(c_n, i)$ pairs for which $r_{ni}$ is known, $\|\ \|$ is the Euclidean norm and $\lambda$ denotes a constant controlling the extent of regularization. In this work, $\lambda = 1.5$.

#### 3.3.4.2 Advanced rating-based collaborative filtering

As a more complex model of this family, we resorted to CF-SVD++, an extension to CF-MF in which the effect of implicit information is included in the minimization rule. The difference here is that the prediction rule considers the fact of a user rating of an item as an additional indication of preference. Therefore, the vector representing the user's interest becomes [26]:

$$p_n + |N(n)|^{-\frac{1}{2}} \sum_{j \in N(n)} y_j \tag{3.16}$$

#### 3.3.4.3 Decision Trees

We first describe the core concept of *Tree-based methods*. The idea consists on splitting the features into a set of nodes or regions, and then fit the data on each node to make the prediction. A popular tree-based model is called *Classification and Regression Tree* (CART for short) [10]

The *Regression tree* version of CART is constructed through binary recursive partitioning, which is an iteration process that split the features into branches. The process continues by splitting each partition into a minimum number of nodes. In order to find the recursive binary splitting, both the splitting variable $X_i$ and a split point z are considered such that the splitting at the split point is:

$$R_1(i,z) = \{X|X_i \geq z\} \textit{ and } R_2(i,z) = \{X|X_i < z\} \tag{3.17}$$

Thus a tree is formally described as:

$$T(X,\Theta) = \sum_{j=1}^{J} \gamma_j I(X \in R_j) \tag{3.18}$$

where $\gamma_j$ constant assign to each terminal node, and $\Theta = \{R_j, \gamma_j\}$. The prediction will be the mean of the response variable in the region or terminal node.

The *classification tree* is similar to the regression tree, except it uses classes to predict the categorical response. A majority vote scheme in the terminal node is finally applied to estimate the prediction.

– ***Classification and Regression problems***

In order to preserve individual preferences, we decided to build a tree for each single user. As a minimum number of observations is required to learn the model, we run experiments with three different thresholds: 26, 30, and 40. They determined the number of total users available on each experiment: 19, 10, and 7 users, respectively, depending on the data in Section 3.2.1.

Two types of prediction problems were considered: regression and classification. In the regression problem the tapa rating is a numeric value and the rating prediction consists on calculating the average of the values on each region of each single learner. Figure 3.7 shows an instance of the regression problem for user number 1377 (u1377) who experienced 54 tapas. The tree consist of a series of splits, the first one made on the character of tapa, the second one on its kind. The tree segments the data into three regions:

$R_1 = \{X|CharacterTapa = Tradition\}$,

$R_2 = \{X|CharacterTapa = Daring, Tapakind = \{Meat\ or\ Sweet\ or\ other\}\}$,

$R_3 = \{X|CharacterTapa = Daring, Tapakind = \{egg\ or\ Fish\ or\ Shellfish\}\}$.

The rating predicted each region is 3, 4, and 4.5, respectively. In the classification problem, the tapa rating is converted into a cathegorical variable in the following way:

$$RatingClass = \begin{cases} H, & \text{if tapaRating=5} \\ M, & \text{if } 3 \leq tapaRating \leq 4 \\ L, & \text{if } tapaRating < 3 \end{cases} \tag{3.19}$$

As in the regression problem, the tree segments the data into three regions $R_1$, $R_2$, and $R_3$. The majority vote is used afterwards to estimate the prediction. Fig. 3.8 pictures the classification tree for user number 1377 as well as the histograms showing the frequency of each class within each region. In this case the prediction is H, M, and M, for each of the three regions.

#### 3.3.4.4 Tree-based ensembles

– ***Boosting Tree***

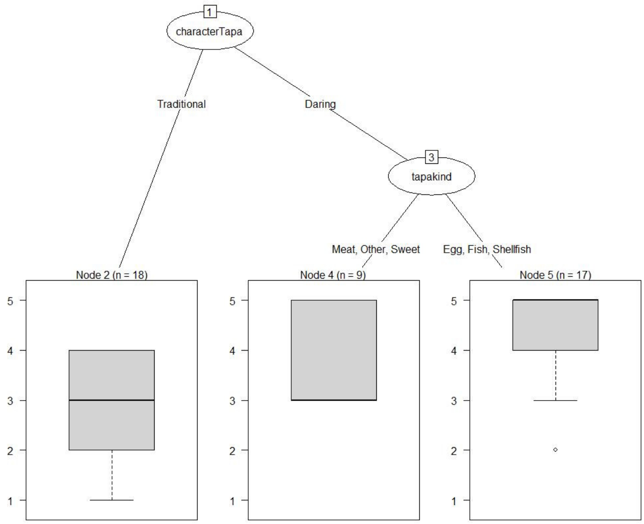


**Figure 3.7:** The regression tree for the user number 1377.

Boosting builds a tree in a sequential fashion. This means each tree depends on the prior trees. Boosting can be implemented in regression as well as classification methods. The boosted tree is:

$$f_m(x) = \sum_{m=1}^{M} T(X, \theta_m) \tag{3.20}$$

where $M$ is the number of trees.

– ***Bagging Tree***

Bagging builds different trees on $M$ different bootstrapped training dataset. All trees are fully grown, and at each node, in the tree one searches over all features to find the feature that best splits the data at that node, and gets the prediction $\hat{f}^{*b}(x)$, and computes the average of prediction in regression:

$$\hat{f}_{bag}(x) = \frac{1}{M} \sum_{m=1}^{M} \hat{f}^{*b}(x) \tag{3.21}$$

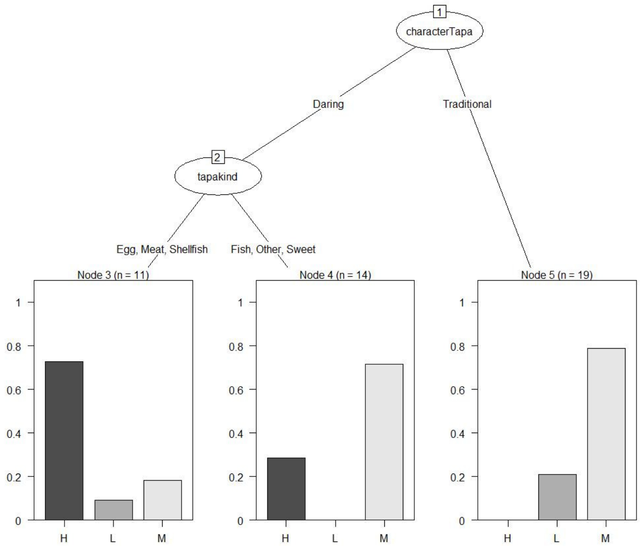


**Figure 3.8:** The classification tree for the user number 1377.

– ***Random Forests*** Random Forests is a particular case of Bagging. The main difference is that at each candidate split in the learning process, a random sample of the predictors or features is chosen among all the predictors or features. The goal is to build a large collection of decorrelated trees. The prediction in a regression problem is estimated as follows:

$$\hat{f}_{rf}(x) = \frac{1}{M}\sum_{m=1}^{M}\hat{f}^{*b}(x) \tag{3.22}$$

In a classification problem, the prediction in Bagging and Random Forests is made by means of the *majority vote* for the classes.

## 3.4 Evaluation

The performance of all models were validated using standard cross-validation procedures. In the case of choice-based models fitted with the Rectur dataset, the outcomes were analyzed by applying random sub-sampling as well as leave-one-out cross-validation. For validation of random sub-sampling, 100 iterations were considered using 25% of randomly selected individuals for testing and the other 75% for training. For each decision-maker in the test data and for each recommendation method, prediction error measures were then estimated. The procedure for leave-one-out cross validation is similar, but the test set includes only one decision-maker per iteration.

In the case of all the models fitted with the MovieStar dataset, a k-fold cross-validation was carried out. Specifically, ten trials of each training-evaluation cycle was performed for every single user. After that, the average of the outcomes for the 10 trials was estimated for each model-user pair. Finally, the results are aggregated for both models and users. For models, the average accuracy over all users is estimated. For users, a performance ranking based on accuracy is obtained.

The following subsections provide a brief description of the performance metrics applied in the evaluation process.

### 3.4.1 Accuracy

The Accuracy measure is estimated as the fraction between the correct and the total number of predictions:

$$Accuracy = \frac{\text{Correct predictions}}{\text{Total predictions}} \tag{3.23}$$

This metrics was used to evaluate all choice-based as well as the classification models used in our experiments.

### 3.4.2 Discounted Cumulative Gain (DCG)

In some experiments, the items were ranked according to the predicted rating or its choice probability. In these cases, the best ranked item is the one predicted to be chosen (Top-1 scheme). In order to measure the ranking quality, the Discounted Cumulative Gain (DCG) was chosen as it captures the distance between the true choice and the predicted choice [24]. This is defined as follows:

$$DCG_p = \sum_{i=1}^{p} \frac{rel_i}{log_2(i+1)}. \tag{3.24}$$

where $p$ is a particular ranking position, and $rel_i$ is the weighted relevance at position $i$. In the present study, there was only one relevant tapas item, i.e. the item chosen, and therefore $rel_i$ was set to 1 when the relevant item was at position i, and 0 otherwise.

### 3.4.3 RMSE

The Root Mean Squared Error (*RMSE*) is defined as the standard deviation of the difference between the actual and the predicted rating.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{r}_i - r_i)^2}{n}} \tag{3.25}$$

where $\hat{r}_i$ being the predicted rating, and $r_i$ the actual one. RMSE was applied to evaluate the regression models used in our experiments.

### 3.4.4 Frequency of Best-Model

The frequency of Best-Model $f_{BestModel}$ is the number of times that a given model is observed in the first position of a ranking that orders the models in terms of their performance for a single user. The performance being measured using the Accuracy metrics described above.

### 3.4.5 MSE

For the informed ensemble method, we used the Mean Square Error (MSE), the mean squared error used to compute a distance between the real and estimated value; this gap called the prediction error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - X_i)^2 \tag{3.26}$$

where $\hat{x}_i$ being the estimation value, and $X_i$ the real value.

### 3.4.6 Crowd Square Error And Diversity

The other prior information used in an informed method the diversity and Crowd Square Error (CSE), the Diversity Prediction Theorem used to measure CSE by the difference between the

MSE and Diversity of Crowd [32]. The diversity estimated as the difference between the average of all prediction and prediction one.

$$AVGPrediction(AVP) = \frac{1}{n}\sum_{i=1}^{n}\hat{x}_i \tag{3.27}$$

$$Diversity = \frac{1}{n}\sum_{i=1}^{n}(\hat{x}_i - AVP)^2 \tag{3.28}$$

The CSE computed through the following equation:

$$CSE = MSE - Diversity \tag{3.29}$$

## 3.5 Statistical Analysis

We have carried out statistical hypothesis tests to confirm whether a difference of average accuracy between two models is statistical significant or not. Because our sample (the number of accuracy values that were used to estimate the average) is small, we resorted to the Student two-sample t-test.

### 3.5.1 Independent two-sample t-test

Hypothesis testing is a statistical approach for testing a hypothesis about a parameter in a population using data measured in a sample. In this testing, we analyze some hypotheses by determining the probability that a sample statistic could have been selected, if the hypothesis including the population parameter were correct. The main goals of statistical hypothesis testing are to estimate the P-value, which is the probability of obtaining the regarded results, or something more extreme if the null hypothesis were correct.

The independent t-test is an inferential statistical test that concludes whether there is a statistically significant difference between the means in two unrelated groups or not. In statistical terms, it means that the researcher is testing the probability that the two groups of data approach from the same population [16].

The null hypothesis for the independent t-test guess the population means from the two unrelated sets are equal:

$$H_0 : \mu_1 = \mu_2$$

To detect if we can reject the null hypothesis and accept the alternative hypothesis, which is that the population means are not equal:

$$H_a : \mu_1 \neq \mu_2$$

In the experiments, we used "Equal sample sizes, unequal variances t-test," is used to test the hypothesis that two populations have equal means. This test is more reliable when the two samples have unequal variances and equal sample sizes. The t-test whether the population means are different is calculated as:

$$t = \frac{\mu_1 - \mu_2}{S\bar{\Delta}} \tag{3.30}$$

where

$$S\bar{\Delta} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{3.31}$$

$S_1^2$ and $S_2^2$ are the unbiased estimator of the variance.

The statistical results are presented in standard form: "*" means $p < 0.05$, "**" means $p < 0.01$, "***" means $p < 0.001$, and "ns" means $p > 0.05$.

## 3.6 Software and tools.

### 3.6.1 Experiments: iMotions

iMotions software is a biometric research platform. Everything is synchronized in real-time with EEG, GRS, FACET and Eye-Tracking sensors in a single unified platform where the obtained data is viewed, analyzed or exported.

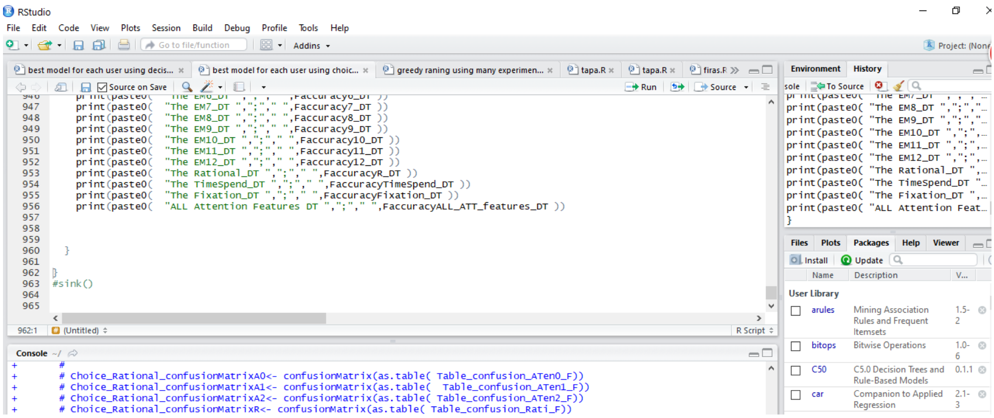


**Figure 3.9:** R Tool Interface.

### 3.6.2 Analysis: R framework

We conduct the analyses in R, the free software environment for statistical computing( see Figure 3.9. Specifically, we used the following packages: (1) the mlogit package to estimate the multinomial logit models (see [12] for further details), (2) the caret package to estimate the single decision-trees and the tree-based ensembles, and (3) the recommenderlab package to evaluate the rating-based baselines. Besides, some cross-validation functions were developed using the R language to analyze the model performance in terms of predictions.

## CHAPTER 4

# RESULTS

## 4.1 Data description

### 4.1.1 Rectur Dataset

The RECTUR data set, which was characterized in Section 3.2.1, included the choices of 5517 individuals regarding a set of 113 tapas available during the Santiago(é)Tapas contest. Although most of these individuals tasted only one tapas item, a large number of users tasted several. The three choice sets corresponding to the three locations of restaurants in the city are briefly described.

In the *new area of the city* 2030 users consumed 3888 tapas that were chosen from 37 alternatives: 18 of traditional character, and 19 of novel character. Figure 4.1 shows a histogram plotting the consumption of each tapas, a label indicating the main ingredient and the average rating per item. According to the data, t22 and t61 were the most popular choices and the average rating was greater than 3, which indicates a high level of satisfaction.

As for the *old town*, 3953 participants tasted 8948 tapas chosen from the set of 62 available tapas: 32 of traditional character, and 30 of novel character. Figures 4.2 and 4.3 show the total number of novel and traditional tapas that users consumed for the 62 possible choices, respectively. According to the data, t101 was the most common choice, while t37, t103 and t102 were rarely selected. With regard to the average, the lowest ratings correspond to t21 and t94, and the highest to t11 and t99.

Finally, in the *outlying area of the city*, the least popular area, 436 users consumed 743 tapas from 14 available choices: 3 of traditional nature and 11 of novel nature. As before, Figure 4.3 summarizes the data set for this area. According to this figure, t44, t45, t104 and t105 were the tapas most frequently chosen, and t2 and t3 were rarely selected. In this case, the lowest and highest mean ratings correspond to tapas t58 and t44, respectively.

### 4.1.2 Movistar Dataset

Data were collected from 39 subjects (20 females, 19 males, age range: 18-51 years, nationality: Spain and Latin-American countries, selection sources: academia and business sectors). All subjects had prior experience using Internet and Web applications and had a normal or corrected-to-normal vision. Due to sampling errors of EEG and FACET devices, some periods of the emotional activity were not properly recorded. This caused the number of valid observations, which may reach up to 20 (one for each choice situation), may be different for each subject.

## 4.2 Baseline ensembles

### 4.2.1 Tree-based ensembles

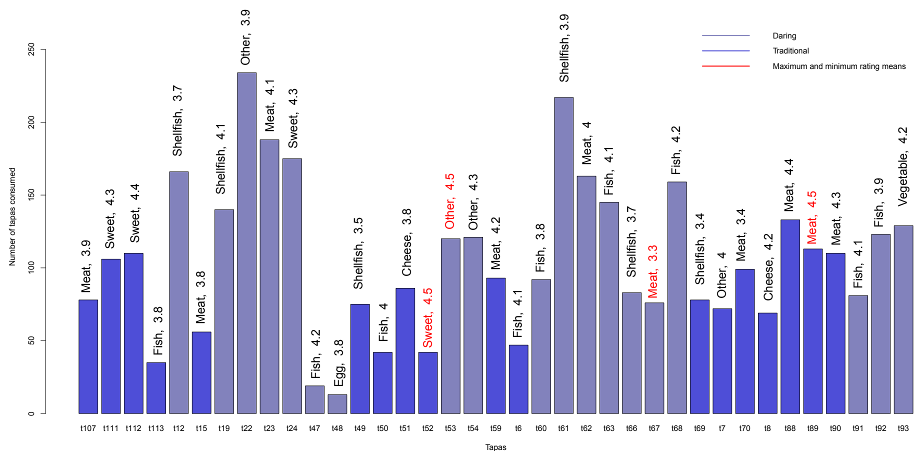


**Figure 4.1:** Bar plot for number of different tapas consumed, main ingredient and mean of users' ratings in the new zone of the city.

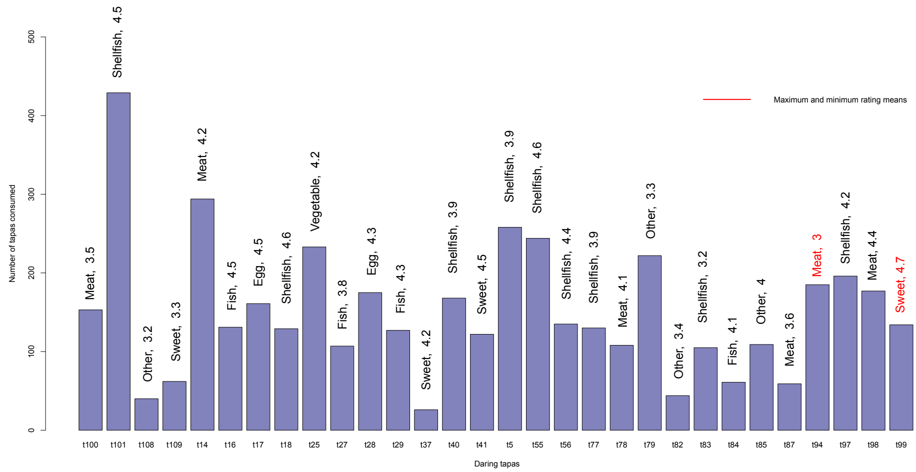


**Figure 4.2:** Bar plot for number of different novel tapas consumed, main ingredient and mean of users' ratings in the old town in the city.

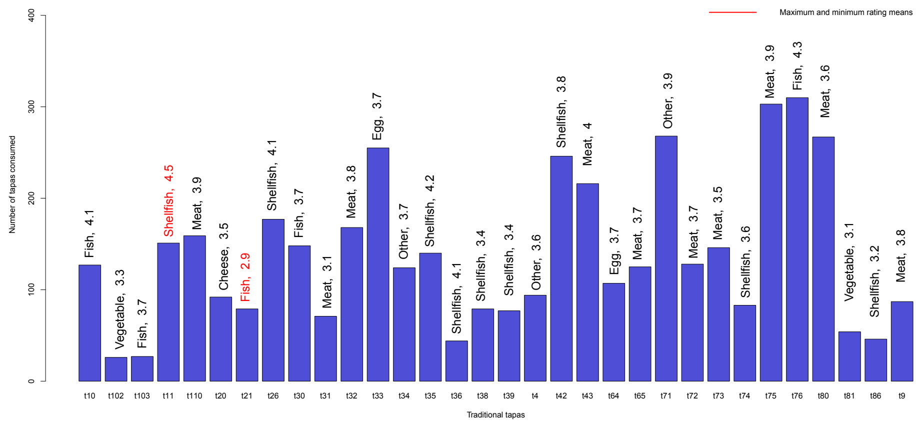


**Figure 4.3:** Bar plot for number of different traditional tapas consumed, main ingredient and mean of users' ratings in the old town.

#### 4.2.1.1 Classification problem

To explore the performance of tree-based ensembles on the classification problem, we first provide some results on individual users. Table 4.1 shows the accuracy for three randomly

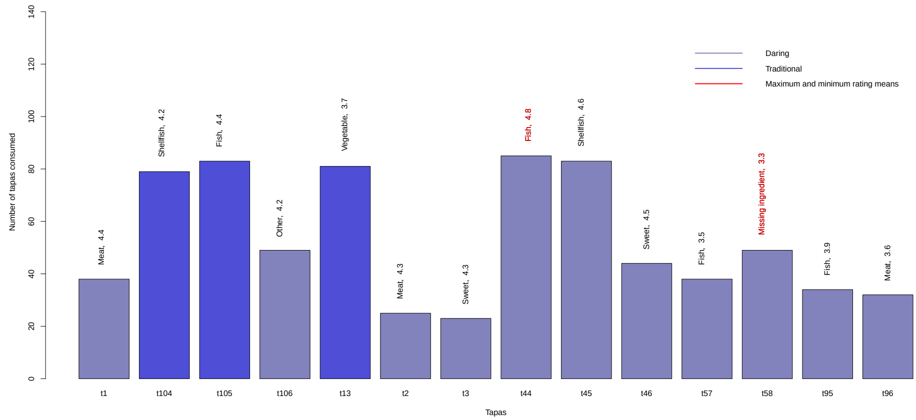


**Figure 4.4:** Bar plot for number of different tapas consumed, main ingredient and mean of users' ratings in the outlying zone of the city.

chosen users satisfying the threshold values that were set in section 3.3.4.3. It can be observed that Random Forest (RF) ranks the highest in all three cases. Afterward, three clusters were created, each cluster made up with those users satisfying the following threshold values: 26, 30, and 40 consumed tapas. The accuracy average on each cluster for each method is shown in Figure 4.5. The numbers confirm that RF outperforms the others in terms of Accuracy in all the clusters.

#### 4.2.1.2 Regression problem

Table 4.2 shows the RMSE for three randomly chosen users satisfying the threshold values that were set in section 3.3.4.3. For this problem, the CART method ranks the highest (lowest RMSE) in all three cases. Figure 4.6 presents the results obtained for the clusters. The plots position GBM, an ensemble method, in the first place of the rank in two of the three clusters (30 and 40), with CART preserving the first place only in cluster with threshold 26.

### 4.2.2 Comparison with CF approaches

The comparison was made for the regression problem only as of the chosen CF approaches (user-based and matrix factorization) work with numeric values rather than classes or cat-

**Table 4.1:** Performance in classification problem: single users. Accuracy results for CART, Treebag, GBM, RF.

| Users | Number of a Tapas consumed | Methods | Accuracy |
|---|---|---|---|
| U1025 | 27 | CART | 0.4 |
| | | Treebag | 0.22 |
| | | GBM | 0.4 |
| | | RF | 0.8 |
| U1377 | 54 | CART | 0.6 |
| | | Treebag | 0.58 |
| | | GBM | 0.5 |
| | | RF | 0.9 |
| U204 | 35 | CART | 0.67 |
| | | Treebag | 0.41 |
| | | GBM | 0.33 |
| | | RF | 1 |

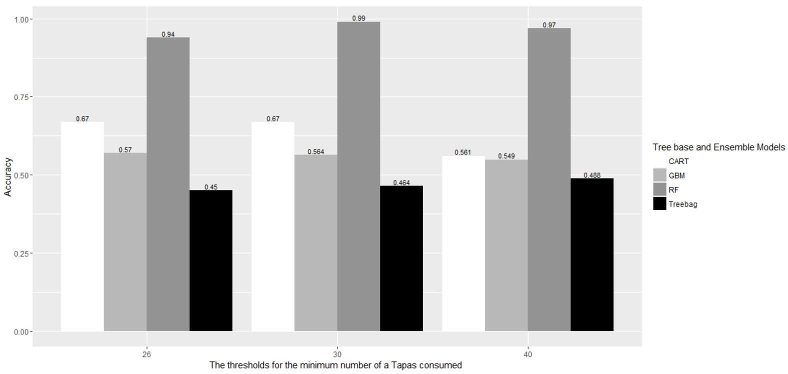


**Figure 4.5:** Performance in classification problem: clusters. Accuracy average for CART, Treebag, GBM, and RF. Results are presented for three cluster of users, each cluster made up with those users satisfying the following threshold values: 26, 30, and 40 consumed tapas.

egories. Figure 4.7 illustrates the results in terms of RMSE for the three clusters already described. The plot shows the superior performance of UBCF, with GBM (clusters 30 and 40) and CART (cluster 26) in the second place. Surprisingly, matrix factorization offers the worse performance of all methods.

## 4.3 Choice-based models

**Table 4.2:** Performance in regression problem: single users. RMSE results for CART, Treebag, GBM, RF.

| Users | Number of a Tapas consumed | Methods | RMSE |
|---|---|---|---|
| U1025 | 27 | CART | 0.98 |
| | | Treebag | 1.56 |
| | | GBM | 1.25 |
| | | RF | 1.18 |
| U1377 | 54 | CART | 0.96 |
| | | Treebag | 1.04 |
| | | GBM | 1.01 |
| | | RF | 1.29 |
| U204 | 35 | CART | 0.94 |
| | | Treebag | 0.99 |
| | | GBM | 1.11 |
| | | RF | 1.03 |

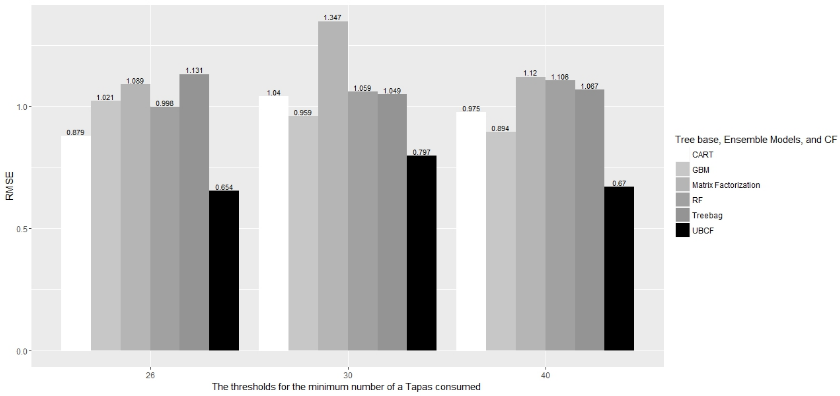


**Figure 4.6:** Performance in regression problem: clusters. RMSE average for CART, Treebag, GBM, and RF. Results are presented for three cluster of users, each cluster made up with those users satisfying the following threshold values: 26, 30, and 40 consumed tapas.

### 4.3.1 Fitting of models

Both the standard and mixed logit models were fitted to the data for the three choice sets described in Section 4.1. For the mixed logit model, a Gaussian distribution of the coefficients was assumed, and the number of draws, $R$, was set to 100.

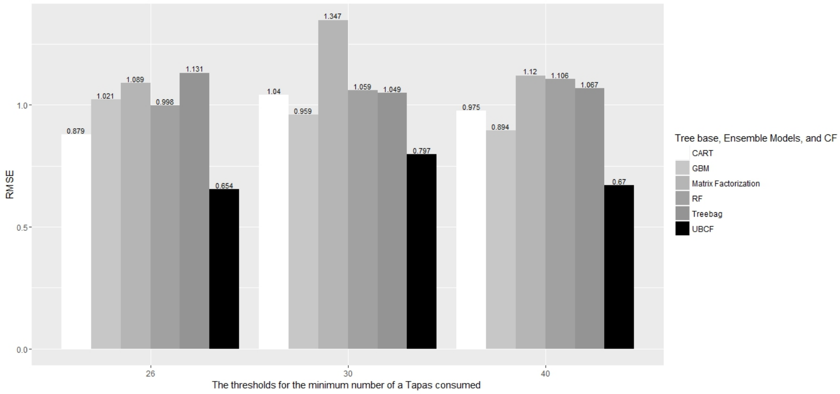


**Figure 4.7:** Performance comparison. RMSE average for CART, Treebag, GBM, RF, UBCF and matrix factorization. Results are presented for three cluster of users, each cluster made up with those users satisfying the following threshold values: 26, 30, and 40 consumed tapas.

The coefficients obtained for both models are shown in Tables 4.3 and 4.4. Most of these proved significant (in black). For the mixed logit model (Table 4.4), only the mean estimations of Gaussian distributions are shown. In terms of preferences, the sign of coefficients represents the positive or negative preference of users for the tapas attribute. For instance, Table 4.3 shows that participants revealed a positive preference for egg, meat, and shellfish tapas in the old town, but a negative one for egg and traditional tapas in the new zone.

| | New zone | Old town | Outlying zone |
|---|---|---|---|
| Cheese | -0.07 | **-0.25** | |
| Egg | **-2.48** | **0.31** | |
| Fish | **-0.46** | -0.02 | 0.14 |
| Meat | 0.06 | **0.28** | **-0.44** |
| Shellfish | -0.03 | **0.21** | **0.38** |
| Sweet | 0.07 | **-0.46** | **-0.38** |
| Vegetable | **-0.18** | **-0.17** | 0.26 |
| Traditional | **-0.62** | **-0.15** | **0.24** |
| Log-Likelihood: | -13772 | -36757 | -1913.8 |

**Table 4.3:** Estimation by maximum likelihood of the standard logit model coefficients for different areas of the city. Significant coefficients are shown in black.

| | New zone | Old town | Outlying zone |
|---|---|---|---|
| Cheese | -0.07 | -0.24 | |
| Egg | **-2.48** | **0.31** | |
| Fish | **-0.46** | -0.01 | 0.13 |
| Meat | -0.07 | **0.27** | **-0.67** |
| Shellfish | -0.03 | **0.21** | **0.37** |
| Sweet | -0.003 | **-0.46** | **-0.38** |
| Vegetable | **-0.18** | **-0.17** | 0.26 |
| Traditional | **-0.93** | **-0.09** | -0.01 |
| Log-Likelihood: | -13631 | -36680 | -1897.9 |

**Table 4.4:** Estimation of the means for mixed logit model coefficients assuming normal distribution for different areas of the city. Significant coefficients are shown in black.

### 4.3.2 Performance evaluation

It is important to point out that only the first choice-based model, the standard logit model, was evaluated and compared with the baseline algorithms. This is because the standard and mixed logit models provide similar estimations for model coefficients according to the results shown in section 4.3.1.

Tables 4.5, 4.6, and 4.7 show the evaluation results for the outlying, new and old areas of the city, respectively. The data shows that choice-based models perform slightly better, in terms of Accuracy, than both CF and Ensemble algorithms, but quite similarly to the Single-Tree approach. However, in most cases, Accuracy is zero or close to zero, indicating that the predicted tapas item does not usually correspond to the one actually chosen. DCG (see section 3.4.2, in turn, is more informative for analyzing and comparing the performance of the different models. Choice-based models showed superior performance in the Outlying area of the city, but Single-Tree and Ensemble algorithms provided better results in the other two areas.

The ranking of the methods in terms of DCG in the R.CV validation is shown in Figure 4.8. In the case of equal DCG values, the ranking was based on Accuracy. Choice-models were ranked first in the outlying area (with only 14 alternative tapas) but the comparative performance was lower in the other two areas, in which the choice sets are larger. Ensembles show a somewhat opposite behaviour: they performed comparatively better when more choices, users and observations are available. Surprisingly, the Single-Tree method, including

| | R.CV | | LOO.CV | |
|---|---|---|---|---|
| Method | Accuracy | DCG | Accuracy | DCG |
| CHOICE | 0.11 | 0,35 | 0,12 | 0,36 |
| CF-UB | 0 | 0,29 | 0 | 0,29 |
| CF-MF | 0 | 0,29 | 0 | 0,29 |
| CF-SVD++ | 0,07 | 0,32 | 0,03 | 0,33 |
| Single-Tree | 0,13 | 0,33 | 0,16 | 0,36 |
| Ensemble-Boosting | 0 | 0,29 | 0 | 0,29 |
| Ensemble-Bagging | 0 | 0,29 | 0 | 0,29 |
| Ensembre-RF | 0 | 0,29 | 0 | 0,29 |

**Table 4.5:** Outlying area of the City: Cross validation predictions errors. Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 14 and DCG measures were estimated according to this ranking size.

| | R.CV | | LOO.CV | |
|---|---|---|---|---|
| Method | Accuracy | DCG | Accuracy | DCG |
| CHOICE | 0.05 | 0,27 | 0,05 | 0,25 |
| CF-UB | 0 | 0,21 | 0 | 0,21 |
| CF-MF | 0 | 0,21 | 0 | 0,21 |
| CF-SVD++ | 0,03 | 0,24 | 0,05 | 0,23 |
| Single-Tree | 0,14 | 0,41 | 0,11 | 0,40 |
| Ensemble-Boosting | 0 | 0,28 | 0 | 0,29 |
| Ensemble-Bagging | 0,07 | 0,40 | 0,05 | 0,39 |
| Ensembre-RF | 0 | 0,29 | 0 | 0,30 |

**Table 4.6:** New area of the City: Cross validation predictions errors. Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 37 and DCG measures were estimated according to this ranking size.

only one learner, provided competitive results in the outlying and new areas. As expected, however, for larger data sets (old part of the city) the performance was not as good as that of ensembles and choice-based models. On the other hand, CF approaches are far from being competitive and occupied the lowest position, except for the outlying area of the city. Another interesting finding is that the performance of all models in terms of DCG was lower as long as the choice set, i.e. the number of available tapas, increased from the outlying area to the old town. This may suggest that the Accuracy problem is probably more complex when the choice set increases and the choice becomes more difficult for the decision-maker.

## 4.4 Cognitive models

| | R.CV | | LOO.CV | |
|---|---|---|---|---|
| Method | Accuracy | DCG | Accuracy | DCG |
| CHOICE | 0.02 | 0,21 | 0,01 | 0,21 |
| CF-UB | 0 | 0,18 | 0 | 0,18 |
| CF-MF | 0 | 0,18 | 0 | 0,18 |
| CF-SVD++ | 0,01 | 0,21 | 0,01 | 0,20 |
| Single-Tree | 0 | 0,21 | 0 | 0,21 |
| Ensemble-Boosting | 0 | 0,22 | 0 | 0,22 |
| Ensemble-Bagging | 0 | 0,20 | 0 | 0,20 |
| Ensembre-RF | 0 | 0,20 | 0 | 0,21 |

**Table 4.7:** Old town: Cross validation predictions errors. Random and leave-one-out cross validation are denoted by R.CV and LOO.CV, respectively. In this area, the number of different tapas offered was 62 and DCG measures were estimated according to this ranking size.

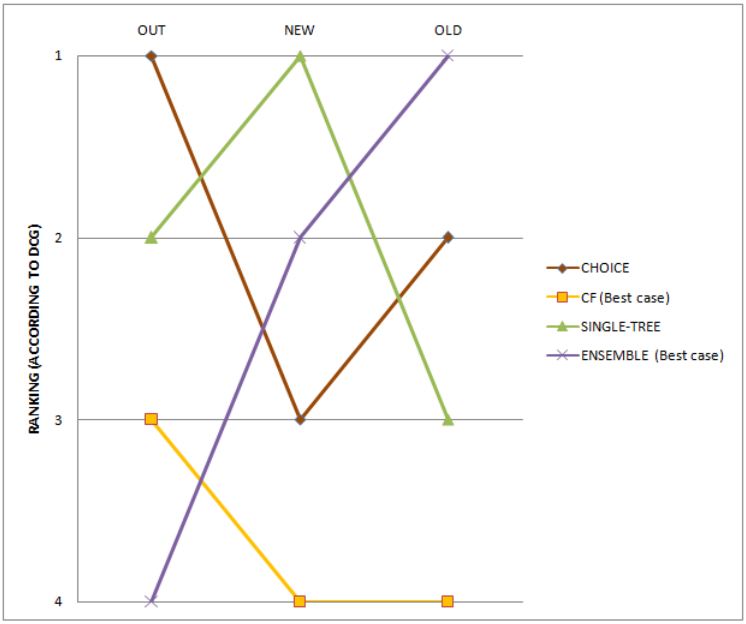


**Figure 4.8:** Ranking of models according to DCG in each area of the city. The performance of models depends on the size of the data set. Choice models work better with fewer available users and fewer alternatives in the choice set (Outlying area of the city), while the ensemble models perform better as the size of the data increases (Old town).

### 4.4.1 Best choice model: all subjects

To find the best model for all subjects we first computed the accuracy of each model for each user, and then estimate the average of all results for all users. (Table 4.8) shows the average

**Table 4.8:** Average performance of single choice models for all users. Accuracy results for all models.

| **Models** | **Average of Accuracy** |
|---|---:|
| A2 | 0.53 |
| A1 | 0.52 |
| R | 0.52 |
| E7 | 0.47 |
| E5 | 0.45 |
| E10 | 0.44 |
| E9 | 0.44 |
| E1 | 0.43 |
| E12 | 0.43 |
| E3 | 0.43 |
| E2 | 0.43 |
| E6 | 0.42 |
| E8 | 0.41 |
| E11 | 0.39 |
| E4 | 0.39 |
| E13 | 0.37 |

of accuracy of all single choice models for all users. It is shown that the first two models in the ranking are the S-I-Attentional models (A2 and A1), which indicate the relevance of the attentional features in the decision-making process. Surprisingly, the S-II-Rational model (R) is located third and all the S-I-Emotional models have ranked afterward. In order to summarize the results, a frequency chart representing the number of times each model type ranked first is shown in Figure 4.9.

### 4.4.2 Best choice model: single subject

To further analyze the performance of the models the results of all single models on each single subject are provided. Based on the Accuracy metrics, Table 4.9 shows the ranking of the Top-5 models for each subject. The key point here is to observe that not only the Top-5 ranking of models but also the best model is different for each subject. For instance, the decision-making process of subject $J02$ seems to follow an emotional-driven process, while subjects $J04$ and $V01$ are better explained with attentional and rational features, respectively.

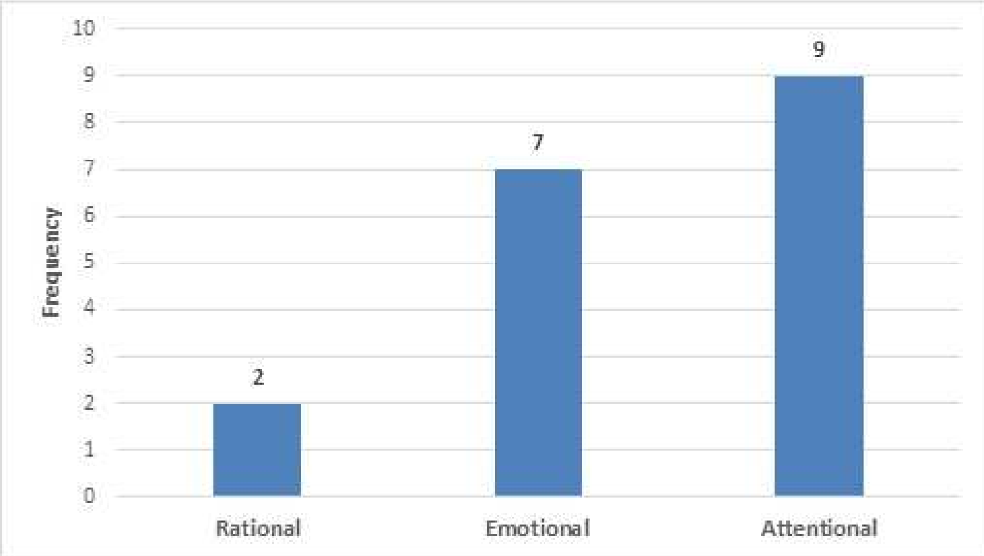


**Figure 4.9:** Frequency of S-II-rational, S-I-emotional, and S-I-attentional models as best model.

This pattern, i.e the best model is different for each subject, is observed for all subjects. Moreover, even the average best model (A2, see Table 4.9) seems to be a bad fit in a significant number of cases. Figure 4.10 shows that the A2 model ranks as lower as places 12, 14, and 15 in the case of 4 subjects.

### 4.4.3 Comparison with baseline ensembles: separated features

In order to compare the performance of single choice models with baseline ensembles, we first carried out a set of experiments where models were fitted splitting the original set of features into three subsets of attentional, emotional, and rational features (see Table 3.5 for further details).

For the attentional subset, the results of all models for each subject are shown in Table 4.10. The single choice model (A) as well as the ensembles (RF-A, Boosting-A, and Bagging-A) were fitted using all attentional features. It can be noticed that the single attentional choice model make a better job than their ensemble's counterparts for most of the

**Table 4.9:** Performance of choice-based models for each subject. Subjects were identified by labels indicating day and experimental slot. Accuracy results for Top-5 models.

| Users | Observations | Top Five Models | Accuracy | Users | Observations | Top Five Models | Accuracy |
|---|---|---|---|---|---|---|---|
| J02 | 12 | E9 | 0.79 | V01 | 15 | R | 0.45 |
| | | E12 | 0.79 | | | E7 | 0.45 |
| | | E5 | 0.76 | | | E11 | 0.45 |
| | | E13 | 0.67 | | | E2 | 0.41 |
| | | E7 | 0.64 | | | E5 | 0.41 |
| J04 | 20 | A1 | 0.64 | V05 | 20 | E5 | 0.66 |
| | | A2 | 0.64 | | | E1 | 0.64 |
| | | E12 | 0.59 | | | E6 | 0.62 |
| | | E9 | 0.58 | | | E7 | 0.62 |
| | | E10 | 0.52 | | | R | 0.62 |
| J06 | 20 | A2 | 0.68 | V08 | 18 | E5 | 0.53 |
| | | A1 | 0.62 | | | R | 0.52 |
| | | E5 | 0.56 | | | E3 | 0.45 |
| | | E10 | 0.56 | | | E6 | 0.44 |
| | | R | 0.55 | | | A2 | 0.42 |
| J07 | 20 | A1 | 0.71 | V09 | 14 | R | 0.52 |
| | | A2 | 0.71 | | | E2 | 0.48 |
| | | R | 0.65 | | | E1 | 0.45 |
| | | E12 | 0.65 | | | E6 | 0.43 |
| | | E7 | 0.61 | | | E7 | 0.43 |
| J10 | 18 | E3 | 0.48 | V13 | 12 | E5 | 0.61 |
| | | E4 | 0.47 | | | R | 0.58 |
| | | E8 | 0.47 | | | E2 | 0.48 |
| | | E2 | 0.42 | | | E7 | 0.48 |
| | | A1 | 0.42 | | | E1 | 0.48 |
| J13 | 20 | A1 | 0.61 | V14 | 16 | A1 | 0.64 |
| | | A2 | 0.59 | | | A2 | 0.59 |
| | | R | 0.52 | | | E3 | 0.48 |
| | | E5 | 0.48 | | | R | 0.47 |
| | | E1 | 0.48 | | | E13 | 0.44 |
| J14 | 12 | E10 | 0.85 | V17 | 20 | A1 | 0.61 |
| | | R | 0.76 | | | A2 | 0.50 |
| | | E7 | 0.58 | | | E7 | 0.48 |
| | | E2 | 0.58 | | | E6 | 0.47 |
| | | E6 | 0.52 | | | E9 | 0.45 |
| J15 | 13 | E12 | 0.48 | V18 | 19 | A1 | 0.76 |
| | | A2 | 0.45 | | | A2 | 0.70 |
| | | A1 | 0.43 | | | R | 0.56 |
| | | E9 | 0.36 | | | E3 | 0.44 |
| | | E11 | 0.36 | | | E5 | 0.44 |
| J19 | 16 | A2 | 0.65 | V21 | 16 | A2 | 0.67 |
| | | A1 | 0.60 | | | R | 0.64 |
| | | E12 | 0.53 | | | A1 | 0.58 |
| | | E1 | 0.49 | | | A7 | 0.55 |
| | | E8 | 0.49 | | | A3 | 0.49 |

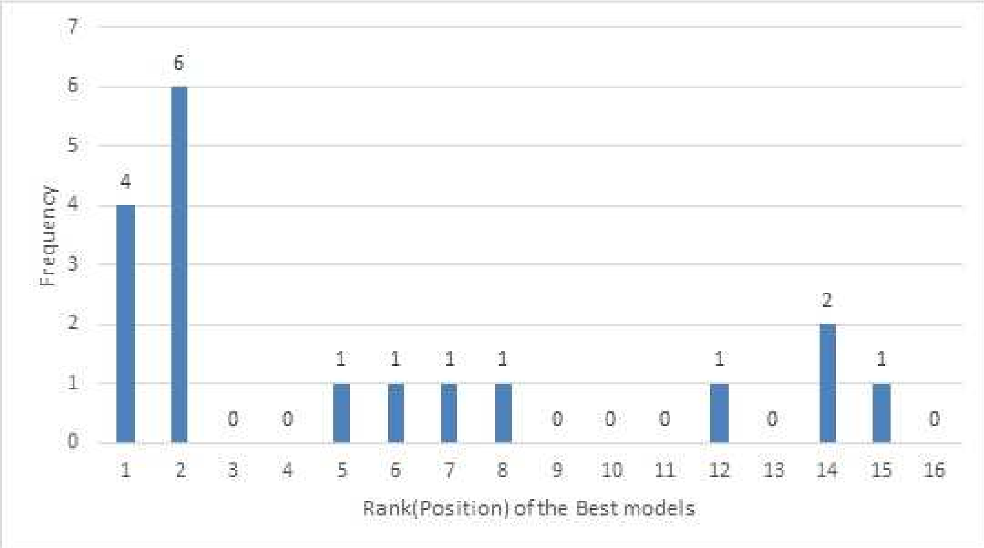


**Figure 4.10:** Rank frequancy of the average best model (A2).

subjects. The comparison is summarized in Table 4.11 which counts the number of subjects in which the ensembles overcome the single models and vice-versa.

For the rational and emotional subsets, the results of all models for each subject are shown in Table 4.12 and Table 4.13. The procedure was the same as followed for the attentional models. The data shows the same trend as observed with attentional models: the single choice models (R and E) perform better than the ensembles built using the same features. Tables 4.15 and 4.14 illustrate the performance comparison for all subjects.

The overall comparison shown in Tables 4.11, 4.14 and 4.15 indicate that ensembles are able to overcome single choice models only in a number of cases.

The quest of the best ensemble for all users is clarified with the figures shown in Table 4.16, where the performance of all ensemble methods using the three separated subsets of features is presented. The results suggest that the best model for each subject not only differs in the subset of preferred features, but also in the ensemble method providing the best prediction.

**Table 4.10:** Performance of single choice model (A) and ensembles (RF-A, Boosting-A, and Bagging-A) for all subjects and attentional features. Performance in terms of accuracy. Best model for each user highlighted in gray.

| **Users, Observations** | **Accuracy Of Models** | | | |
|---|---|---|---|---|
| | **A** | **RF-A** | **Boosting-A** | **Bagging-A** |
| J04, 20 | 0.60 | 0.57 | **0.67** | 0.60 |
| J06, 20 | **0.67** | 0.38 | 0.33 | 0.33 |
| J07, 20 | **0.59** | 0.48 | 0.48 | 0.41 |
| J10, 18 | 0.29 | **0.38** | 0.29 | 0.33 |
| J13, 20 | **0.74** | 0.44 | 0.37 | 0.48 |
| J19, 16 | **0.57** | 0.48 | 0.33 | 0.24 |
| V01, 15 | 0.38 | 0.52 | 0.43 | **0.57** |
| V05, 20 | **0.48** | 0.33 | **0.48** | 0.37 |
| V08, 18 | **0.50** | 0.33 | 0.38 | 0.38 |
| V14, 16 | **0.56** | 0.50 | **0.56** | 0.44 |
| V17, 20 | **0.59** | 0.56 | 0.44 | 0.41 |
| V18, 19 | **0.59** | 0.19 | 0.15 | 0.15 |
| V21, 16 | 0.56 | **0.72** | 0.61 | 0.67 |

**Table 4.11:** Performance comparison between single choice model (A) and the best ensemble (RF-A, Boosting-A, or Bagging-A).

| **Comparison case** | **Frequency** |
|---|---:|
| Ensembles > Single model | 4 |
| Ensembles = Single model | 2 |
| Ensembles < Single model | **7** |

**Table 4.12:** Performance of single choice model (R) and ensembles (RF-R, Boosting-R, and Bagging-R) for all subjects and rational features. Performance in terms of accuracy. Best model for each user highlighted in gray.

| Users, Observations | Accuracy Of Models | | | |
|---|---|---|---|---|
| | R | RF-R | Boosting-R | Bagging-R |
| J04, 20 | **0.53** | 0.33 | 0.43 | 0.33 |
| J06, 20 | **0.54** | 0.46 | 0.46 | 0.29 |
| J07, 20 | **0.48** | 0.33 | 0.22 | 0.41 |
| J10, 18 | 0.29 | 0.38 | 0.50 | **0.58** |
| J13, 20 | 0.41 | **0.48** | 0.30 | 0.41 |
| J19, 16 | 0.52 | 0.48 | 0.38 | **0.57** |
| V01, 15 | 0.38 | 0.43 | 0.52 | **0.62** |
| V05, 20 | **0.59** | 0.37 | 0.33 | 0.33 |
| V08, 18 | **0.54** | 0.46 | 0.33 | 0.50 |
| V14, 16 | **0.33** | 0.22 | 0.17 | 0.11 |
| V17, 20 | **0.30** | 0.26 | 0.22 | 0.19 |
| V18, 19 | **0.52** | 0.37 | 0.44 | 0.26 |
| V21, 16 | 0.56 | **0.67** | 0.50 | 0.56 |

### 4.4.4 Comparison with baseline ensembles: aggregated features

The experimental evidence presented in the previous section showed that ensembles built with separated features could not overcome single choice models. So, what about building ensembles by aggregating all the features (attentional, emotional, and rational)? The last set of experiments compared those ensembles with their single choice models counterparts. The results shown in table 4.17 suggest that ensembles built this way generate more accurate predictions than ensembles built with a separated set of features. Specifically, Boosting-A-R-E seems to be as competitive as the single choice model. In summary, by checking the data presented in table 4.18, it is confirmed that the performance of the best of all three ensembles is slightly better as the single choice model.

**Table 4.13:** Performance of single choice model (E) and ensembles (RF-E, Boosting-E, and Bagging-E) for all subjects and emotional features. Performance in terms of accuracy. Best model for each user highlighted in gray..

| Users, Observations | Accuracy Of Models | | | |
|---|---|---|---|---|
| | E | RF-E | Booting-E | Bagging-E |
| J04, 20 | **0.56** | 0.22 | 0.17 | 0.17 |
| J06, 20 | 0.33 | 0.33 | **0.39** | 0.11 |
| J07, 20 | 0.39 | 0.33 | 0.39 | **0.50** |
| J10, 18 | **0.61** | 0.44 | 0.44 | 0.44 |
| J13, 20 | **0.28** | **0.28** | **0.28** | 0.22 |
| J19, 16 | **0.67** | 0.13 | 0.33 | 0.20 |
| V01, 15 | **0.33** | 0.17 | 0.08 | 0.17 |
| V05, 20 | **0.71** | 0.29 | 0.24 | 0.24 |
| V08, 18 | **0.28** | 0.17 | 0.17 | 0.11 |
| V14, 16 | **0.33** | 0.22 | **0.33** | 0.17 |
| V17, 20 | **0.50** | 0.44 | 0.33 | 0.39 |
| V18, 19 | **0.39** | 0.33 | 0.33 | **0.39** |
| V21, 16 | 0.33 | **0.53** | 0.40 | 0.33 |

**Table 4.14:** Performance comparison between single choice model (E) and the best ensemble (RF-E, Boosting-E, or Bagging-E).

| Comparison case | Frequency |
|---|---|
| Ensembles > Single mode | 4 |
| Ensembles = Single mode | 2 |
| Ensembles < Single mode | **7** |

**Table 4.15:** The Table is Show the Number of Cases Where Ensembles Overcome Single Models and Vice-Versa Using Rational (R) Model.

| Comparison case | Frequency |
|---|---|
| Ensembles > Single mode | 5 |
| Ensembles = Single mode | 0 |
| Ensembles < Single mode | **8** |

**Table 4.16:** Performance comparison of all three ensembles methods (RF, Boosting, Bagging) using the three separated subsets of features (A, E, and R).

| Users | Ensembles Methods | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | RF | | | Boosting | | | Bagging | | |
| | A | E | R | A | E | R | A | E | R |
| J04 | **0.57** | 0.22 | 0.33 | **0.67** | 0.17 | 0.43 | 0.22 | 0.17 | **0.33** |
| J06 | 0.38 | 0.33 | **0.46** | 0.33 | 0.39 | **0.46** | **0.33** | 0.11 | 0.29 |
| J07 | **0.48** | 0.33 | 0.33 | **0.48** | 0.39 | 0.22 | 0.33 | **0.50** | 0.41 |
| J10 | 0.38 | **0.44** | 0.38 | 0.29 | 0.44 | **0.50** | 0.44 | 0.44 | **0.58** |
| J13 | 0.44 | 0.28 | **0.48** | **0.37** | 0.28 | 0.30 | 0.28 | 0.22 | **0.41** |
| J19 | **0.48** | 0.13 | **0.48** | 0.33 | 0.33 | **0.38** | 0.13 | 0.20 | **0.57** |
| V01 | **0.52** | 0.17 | 0.43 | 0.43 | 0.08 | **0.52** | 0.17 | 0.17 | **0.62** |
| V05 | 0.33 | 0.29 | **0.37** | **0.48** | 0.24 | 0.33 | 0.29 | 0.24 | **0.33** |
| V08 | 0.33 | 0.17 | **0.46** | **0.38** | 0.17 | 0.33 | 0.17 | 0.11 | **0.50** |
| V14 | **0.50** | 0.22 | 0.22 | **0.56** | 0.33 | 0.17 | **0.22** | 0.17 | 0.11 |
| V17 | **0.56** | 0.44 | 0.26 | **0.44** | 0.33 | 0.22 | **0.44** | 0.39 | 0.19 |
| V18 | 0.19 | 0.33 | **0.37** | 0.15 | 0.33 | **0.44** | 0.33 | **0.39** | 0.26 |
| V21 | **0.72** | 0.53 | 0.67 | **0.61** | 0.40 | 0.50 | 0.53 | 0.33 | **0.56** |

With regard to the best ensemble for all users, the figures for each subject shown in table 4.17 suggest there is no such thing as the best ensemble. For instance, while subject J04 is best explained with Boosting-A-R-E, subject J06 seems to prefer RF-A-R-E.

## 4.5 Choice-based ensembles

### 4.5.1 Blind methods

The results show the following comparisons: among 1-Learner ensembles, among N-Learner ensembles, and between 1-Learner and N-Learners approaches.

**Table 4.17:** Performance of single choice model (A) and ensembles (RF-A-R-E, Boosting-A-R-E, and Bagging-A-R-E) for all subjects and attentional features. Performance in terms of accuracy. Best model for each user highlighted in gray.

| Users, Observations | Accuracy Of Models | | | |
|---|---|---|---|---|
| | **A-R-E** | **RF-A-R-E** | **Boosting-A-R-E** | **Bagging-A-R-E** |
| J04, 20 | 0.50 | 0.44 | **0.56** | 0.33 |
| J06, 20 | **0.50** | 0.33 | 0.22 | 0.28 |
| J07, 20 | **0.61** | 0.39 | 0.28 | 0.28 |
| J10, 18 | 0.22 | 0.61 | **0.67** | 0.61 |
| J13, 20 | **0.67** | 0.44 | 0.39 | 0.44 |
| J19, 16 | **0.80** | 0.33 | 0.27 | 0.33 |
| V01, 15 | 0.17 | **0.42** | **0.42** | **0.42** |
| V05, 20 | **0.38** | 0.29 | 0.24 | 0.29 |
| V08, 18 | 0.22 | 0.28 | 0.22 | **0.33** |
| V14, 16 | **0.50** | **0.50** | 0.33 | 0.33 |
| V17, 20 | 0.44 | **0.50** | **0.50** | **0.50** |
| V18, 19 | 0.28 | 0.33 | **0.39** | 0.17 |
| V21, 16 | 0.40 | **0.73** | 0.53 | 0.67 |

**Table 4.18:** Performance comparison between single choice model (A-R-E) and the best ensemble (RF-A-R-E, Boosting-A-R-E, or Bagging-A-R-E).

| **Comparison case** | **Frequency** |
|---|---|
| Ensembles > Single mode | **6** |
| Ensembles = Single mode | 2 |
| Ensembles < Single mode | 5 |

– ***Accuracy performance***

For 1-Learner ensembles, the two aggregation techniques were compared in terms of the accuracy of predictions. Table 4.19 shows the results for emotional, attentional, and rational choice-based ensembles. The figures illustrate that performance is quite similar on using the two schemes. As a result, we chose the probability average (PA) as the aggregation techniques for the rest of experiments carried out in this work. In addition, we tested the performance of bagging and boosting sampling strategies. The comparison is shown in table 4.20 points out that the bagging strategy outperforms the boosting one in all three models. The difference is significant in all cases, so bagging

**Table 4.19:** 1-Learner Ensembles: comparison of aggregation techniques: Probability Average and Majority Vote. Three choice-based ensembles were built using the bagging sampling method. The average of accuracy over all users is shown.

| **Models** | Aggregation Technique: **Probability Average** | Aggregation Technique: **Majority Vote** |
|---|---|---|
| Emotion-Choice-Bagging | 0.38 | 0.37 |
| Attention-Choice-Bagging | 0.55 | 0.55 |
| Rational-Choice-Bagging | 0.46 | 0.46 |

**Table 4.20:** 1-Learner Ensembles: comparison of sampling strategies: bagging and boosting. Three choice-based ensembles were built using the probability average (PA) as the aggregation technique. The average of accuracy over all users is shown.

| **Models** | Sample Strategy: **Bagging** | Sample Strategy: **Boosting** |
|---|---|---|
| Emotion-Choice-PA | **0.38**(***) | 0.25 |
| Attention-Choice-PA | **0.55**(***) | 0.34 |
| Rational-Choice-PA | **0.46**(***) | 0.26 |

seems to be the best method to build 1-Learner ensembles. Among the bagging models, the Attention-Choice-PA-Bagging is the best one and offers better performance than the Rational-Choice-PA-Bagging (two-samples t-test, $p < 0.05$).

For N-learner ensembles, the comparison to be made is between common and personalized strategies. In order to test the trade-off between the best performance of top learners and the diversity provided by the full set of all learners, different Top-N ensembles were built and evaluated. Table 4.21 presents the results. Two conclusions can be drawn from them: (1) the personalized strategy achieves the best performance for all Top-N instances, and (2) the addition of learners with lower performance (Top-5, Top-10, and All arrangements) reduces the accuracy of the ensemble. It seems in this case that diversity provided by additional weak learners is not enough to compensate for the loss of overall accuracy.

Finally, a comparison between 1-Learner and N-Learners ensembles needs to be made to assess the best method. The results of the experiments carried out with the best arrangements of each type, shown in table 4.22, confirm the superior performance of N-learners ensembles. Only the A-Choice-Bagging presents a performance in range with the ones of the N-Learners group.

**Table 4.21:** N-Learners Ensembles: comparison of personalized and common strategies. The PA aggregation technique was used in all cases. The average accuracy for Top-2, Top-5, Top-10, and Full ensembles is shown.

| **Learners in ensemble** | Strategy: **Personalized** | Strategy: **Common** |
|---|---|---|
| Top-2 | **0.61**(ns) | 0.58 |
| Top-5 | **0.59**(ns) | 0.54 |
| Top-10 | **0.52**(ns) | 0.51 |
| All | 0.48 | 0.48 |

**Table 4.22:** Comparison between 1-Learner and N-learners ensembles. The average accuracy is shown and compared for all cases.

| **1-Learner** | **N-Learners Personalized:** Top-2 | **N-Learners Common** Top-2 |
|---|---|---|
| E-Choice-Bagging | 0.38-**0.61**(**) | 0.38-**0.58**(**) |
| R-Choice-Bagging | 0.46-**0.61**(**) | 0.46-**0.58**(**) |
| A-Choice-Bagging | 0.55-**0.61**(ns) | 0.55-**0.58**(ns) |

– ***Frequency-Best-Model performance***

The evaluation of models by means of $f_{BestModel}$ provides a supplementary perspective of the behavior of models. First, the ranking of 1-Learner ensembles (shown in Figure. 4.11) confirms that bagging outperforms boosting sampling strategies. Second, and more interesting, it illustrates that Emotion-Choice-Bagging achieves a better result than Attention-Choice-Bagging. This evidence is explained by the fact that whereas Emotion-Choice-Bagging is a specially suitable predictor for a significant number of cases (13 subjects), it provides rather poor predictions for everyone else (5 subjects).

With regard to N-learners ensembles, Figure. 4.12 shows similar performance for both personalized and common strategies. Surprisingly, the comparison between 1-Learner and N-Learners ensembles presents a superior performance of the former models (Figure. 4.13). The rationale of these results, which seem contradictory to the ones obtained with the accuracy metrics, maybe similar to the one provided in the previous paragraph. Models with good average accuracy do not guarantee the best performance for the majority of subjects.

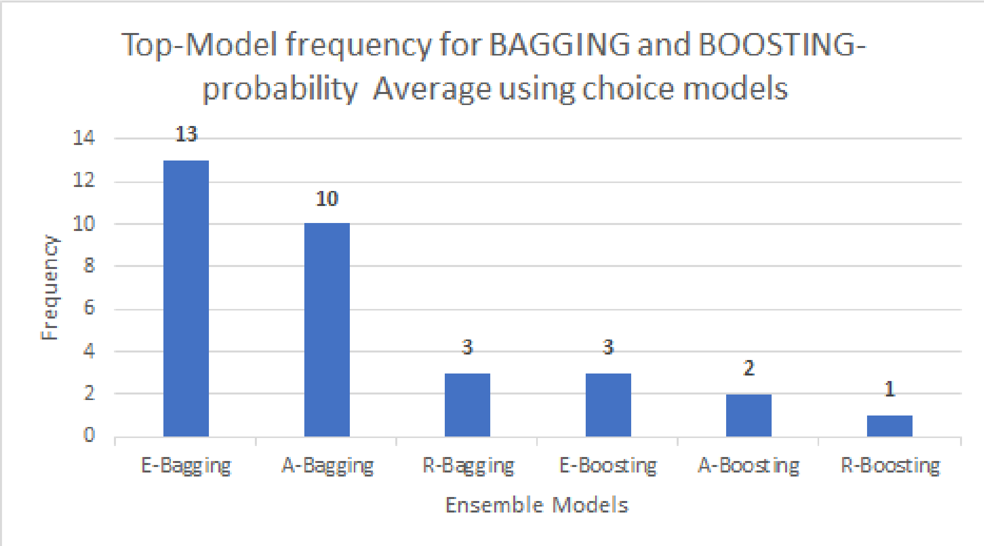


**Figure 4.11:** 1-Learner Ensembles: comparison of sampling strategies in terms of $f_{BestModel}$.

– ***Comparison with single learners and random predictions***

– ***Accuracy performance***

As for 1-Learner ensembles, three choice-based ensembles were built using the bagging sampling method as well as the probability average (PA) as the aggregation technique. Table 4.23 illustrates their performance when compared with their corresponding single choice models. The results are slightly better in two cases (emotional and attentional) but worse in the case of the rational model. The figures illustrate that single choice models seems to be strong rather than weak learners. This probably explains the superior performance of single models against boosting-based ensembles (See table 4.20). As for N-Learners ensembles, in order to apply the same model for all subjects, only the common strategy was considered. Their performance is compared with the Attentional single choice-based model, the one with the best accuracy performance of all models. It is observed that the N-Learner approach provides better results in both Top-2 and Top-2 arrangements (Table 4.24). Finally, it is confirmed that all tested ensemble models worked better than random prediction.

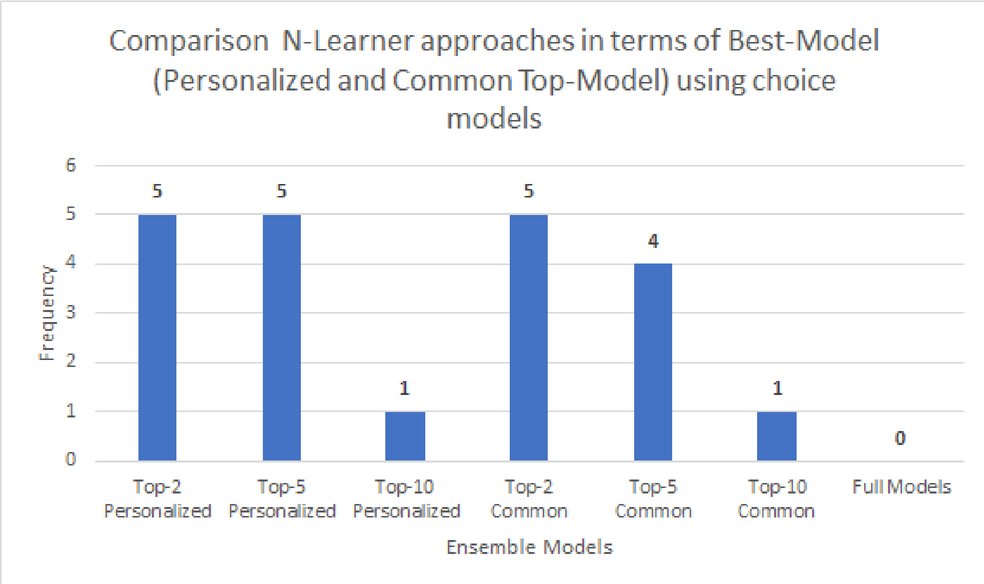


**Figure 4.12:** N-Learners Ensembles: comparison of common and personalized strategies in terms of $f_{BestModel}$.

**Table 4.23:** Comparison of 1-Learner ensembles with single choice models. The average of accuracy over all users is shown.

| **Models** | 1- Learner Ensemble: Bagging-PA | Single Learner |
|---|---|---|
| Emotion-Choice | **0.38**(ns) | 0.35 |
| Rational-Choice | 0.46 | **0.47**(ns) |
| Attentional-Choice | **0.55**(ns) | 0.52 |
| Random Prediction | 0.21 | |

– ***Frequency-Best-Model performance***

The comparison between 1-Learner ensembles and single models in terms of $f_{BestModel}$ is in agreement with what it was found by using the accuracy metrics: 1-Learner models built using bagging outperform their corresponding single models in two (Emotional and Attentional) out of the three cases. Moreover, the plots illustrate the poor performance of the boosting sampling strategy. In all cases, the single models offer better results than the boosting-based ensembles. With regard to N-Learner ensembles, the common strategy outperforms the single model

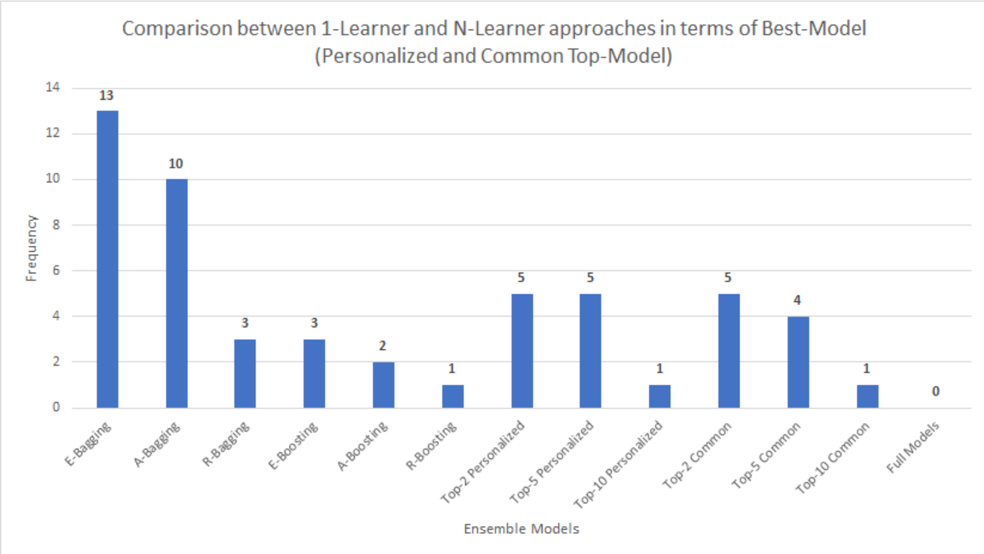


**Figure 4.13:** Comparison between the best 1-Learner and N-learners ensembles in terms of $f_{BestModel}$.

**Table 4.24:** Comparison of N-Learners (Common Top-2 and Top-5) ensembles with single choice models. The average of accuracy over all users is shown.

| **Single Learner** | **N-Learner Common:** Top-2 | **N-Learner Common:** Top-5 |
|---|---|---|
| Emotion-Choice | 0.35-**0.58**(***) | 0.35-**0.54**(***) |
| Rational-Choice | 0.49-**0.58**(***) | 0.49-**0.54**(***) |
| Attentional-Choice | 0.53-**0.58**(*) | 0.53-**0.54**(ns) |
| Random Prediction | 0.21 | |

only under the Top-2 arrangement. When the aggregation includes weaker models (Top-2, Top-10, and All), the generated predictions get worse and single models find their way to the top of the ranking (see Figure.4.14).

- ***Comparison with blind ensembles built with Decision Trees***

  - ***Accuracy performance***

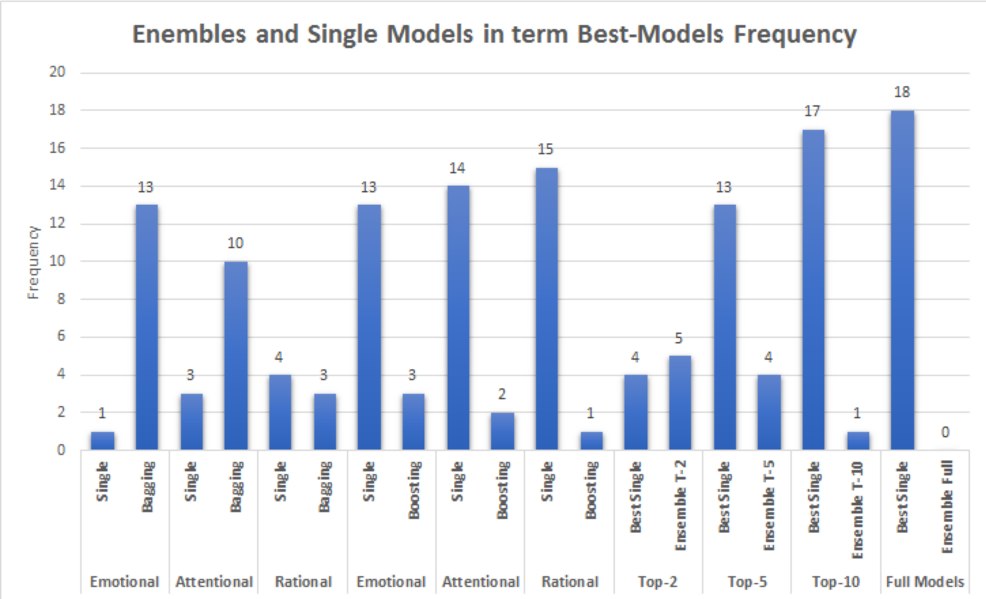


**Figure 4.14:** Comparison between ensembles (1-learner and Common N-learner) and single choice-based models in terms of $f_{BestModel}$.

Three 1-Learner choice-based ensembles were built using the bagging sampling method, and the other three using the boosting one. Both implemented the probability average (PA) as an aggregation technique. They were compared with ensembles built with decision-trees as single learners and the majority vote as aggregation technique (decision trees do not estimate probabilities). Table 4.25 shows the superior performance of choice-based ensembles using the bagging method. However, DT-based ensembles offer better results when compared under the boosting approach. Boosting seems to deepen the performance of single choice models, which seems to confirm their classification as strong single learners.

As for N-Learners, Table 4.26 shows the results for the common strategy. The Top-5 models in terms of average accuracy for choice-based as well as DT-based approaches are listed. In both cases, the attentional features provide the basis of the best single models.

It is observed that the set of Top-5 choice-based models offer higher accuracy values than the corresponding Top-5 DT-based models. The performance of single models also has an impact on the behavior of N-Learners ensembles: the Common-Top-5-Choice clearly outperforms the Common-Top-5-DT. The same pattern is shown with regard the personalized strategy. Table 4.27 presents the performance for 4 randomly chosen subjects. In all cases, the Personalized-Top-5-Choice shows higher accuracy than their corresponding DT counterparts. The average accuracy for all users also confirms the superior performance of choice-based ensembles for all Top-N arrangements (see Table 4.28).

– ***Frequency-Best-Model performance***

The comparison of choice-based and DT-based ensembles under both 1-Learner and N-Learners arrangements is shown in Fig. 4.15. When analyzing the results of the 1-Learner type, it is observed that DT-based ensembles do a better job than their choice-based counterparts, the figures been even better for the boosting sampling strategy. As for the N-Learners type, the observed $f_{BestModel}$ are quite

**Table 4.25:** Comparison between learner types in 1-Learner ensembles. The performance of choice-based and DT-based ensembles in terms of average accuracy over all users is shown.

| **Models** | Learner: **Choice** | Learner: **DT** |
|---|---|---|
| Emotion-Bagging | **0.38**(ns) | 0.34 |
| Rational-Bagging | **0.46**(**) | 0.34 |
| Attention-Bagging | **0.55**(***) | 0.39 |
| Emotion-Boosting | 0.25 | **0.33**(*) |
| Rational-Boosting | 0.26 | **0.33**(ns) |
| Attention-Boosting | 0.34 | **0.34**(ns) |

**Table 4.26:** N-Learners Ensembles: comparison of personalized and common strategies. The PA aggregation technique was used in all cases. The average accuracy for Top-2, Top-5, Top-10, and Full ensembles is shown.

| **Single and Ensemble Models** | Learner: **Choice** | Learner: **DT** |
|---|---|---|
| A1 | **0.54** | 0.40 |
| A2 | **0.51** | 0.39 |
| R | **0.50** | 0.37 |
| E5 | **0.47** | 0.37 |
| E6 | **0.47** | 0.35 |
| Common-Top-5 | **0.54**(***) | 0.40 |

**Table 4.27:** Comparison between learner types in N-Learners ensembles built with the personalized strategy. The performance of both single and ensemble models for a set of subjects (J02, J06, V09 and V21) in terms of accuracy is shown.

| **J02** | | **J06** | |
|---|---|---|---|
| **Single and Ensemble Models** | **Accuracy** | **Single and Ensemble Models** | **Accuracy** |
| E5- Choice | 0.43 | E5-Choice | 0.55 |
| E7-Choice | 0.5 | E10-Choice | 0.53 |
| E9-Choice | 0.93 | R- Choice | 0.53 |
| E12-Choice | 0.77 | A1- Choice | 0.65 |
| E13-Choice | 0.6 | A2-Choice | 0.7 |
| **Personalized-Top-5-Choice** | 0.8 | **Personalized-Top-5-Choice** | 0.72 |
| E3-DT | 0.43 | E3-DT | 0.3 |
| E4-DT | 0.47 | E4-DT | 0.3 |
| E5-DT | 0.5 | E10-DT | 0.3 |
| E6-DT | 0.4 | E11-DT | 0.33 |
| R-DT | 0.43 | A1-DT | 0.32 |
| **Personalized-Top-5-DT** | 0.47 | **Personalized-Top-5-DT** | 0.3 |
| **V09** | | **V21** | |
| **Single and Ensemble Models** | **Accuracy** | **Single and Ensemble Models** | **Accuracy** |
| E1- Choice | 0.43 | E7-Choice | 0.54 |
| E2-Choice | 0.5 | E3-Choice | 0.38 |
| E6- Choice | 0.4 | A1-Choice | 0.58 |
| E7-Choice | 0.45 | A2- Choice | 0.66 |
| R- Choice | 0.58 | R- Choice | 0.64 |
| **Personalized-Top-5-Choice** | 0.58 | **Personalized-Top-5-Choice** | 0.68 |
| E3-DT | 0.43 | E4-DT | 0.6 |
| E4-DT | 0.43 | E3-DT | 0.6 |
| E8-DT | 0.43 | A1-DT | 0.58 |
| E10-DT | 0.45 | A2-DT | 0.62 |
| E11- DT | 0.43 | E6-DT | 0.6 |
| **EPersonalized-Top-5-DT** | 0.43 | **Personalized-Top-5-DT** | 0.62 |

similar for the ensembles built with both learner types. In summary, it turns out that the peak performance of DT-based ensembles, i.e the best outcomes of these predictors for some individual subjects, is comparable to the one obtained through choice-based ensembles.

**Table 4.28:** Comparison between learner types in N-Learners ensembles built with both common and personalized strategy. The performance of choice-based and DT-based ensembles in terms of average accuracy over all users is shown.

| | **Common Strategy** | | **Personalized Strategy** | |
|---|---|---|---|---|
| **Ensemble Models** | Learner: **Choice** | Learner: **DT** | Learner: **Choice** | Learner: **DT** |
| **TOP-2** | **0.58**(**) | 0.44 | **0.61**(***) | 0.42 |
| **TOP-5** | **0.54**(**) | 0.40 | **0.59**(***) | 0.41 |
| **TOP-10** | 0.51 | 0.37 | 0.52 | 0.40 |
| **All** | 0.48 | 0.39 | 0.48 | 0.39 |

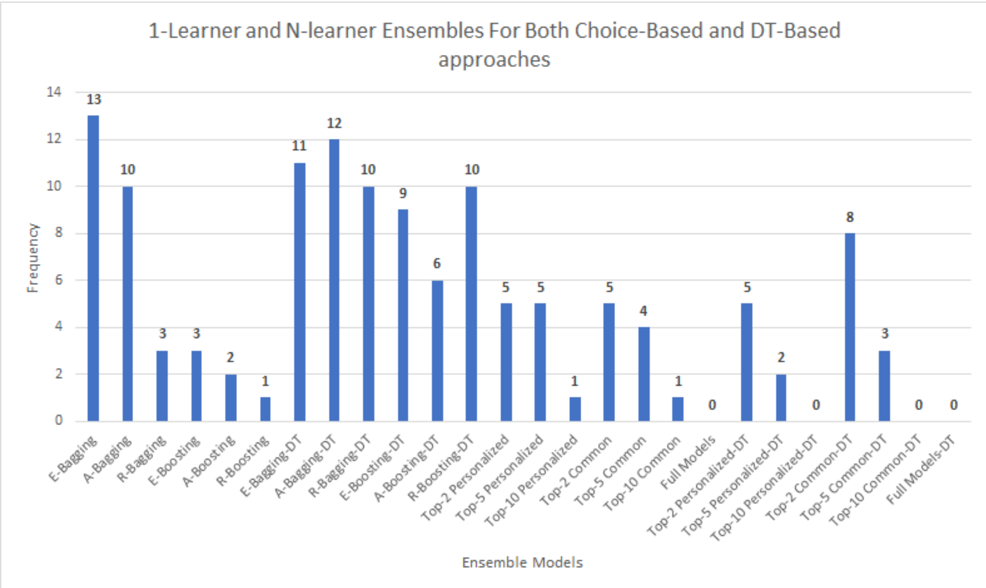


**Figure 4.15:** Comparison between 1-Learner and N-learner Ensembles For Both Choice-Based and DT-Based approaches in terms of Best Models (ranking).

### 4.5.2 Informed methods

The results show the following comparisons among informed ensembles model based on different types of prior information :(1) High diversity, (2) Low error prediction(MSE), (3), and Low crowd error. And between common and personalized strategies.

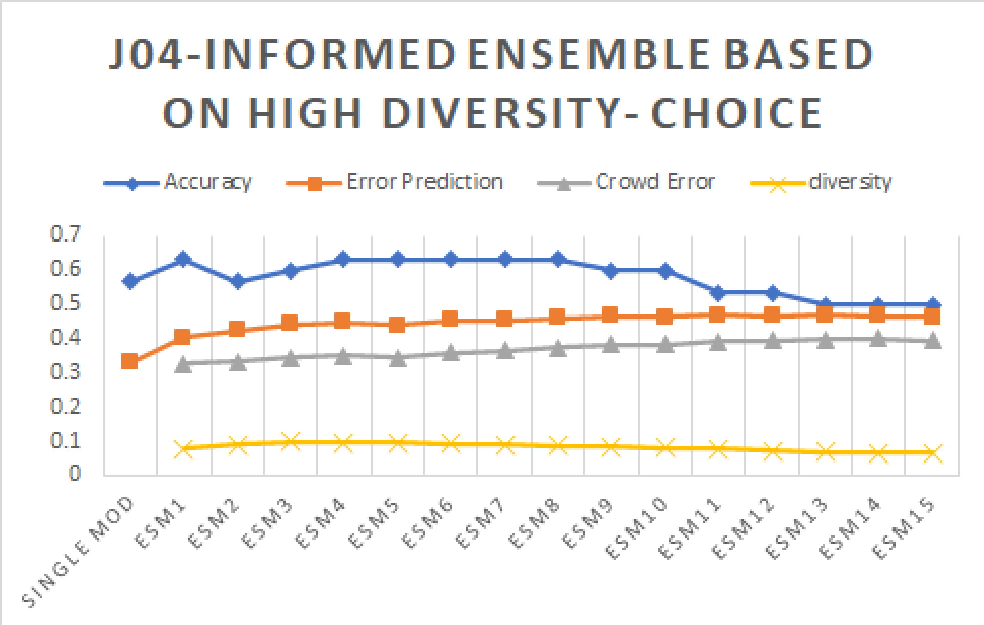


**Figure 4.16:** Informed Ensemble using Greedy Ensemble selection algorithm Based on High Diversity for the subject J04. ESM is mean the Ensemble Models.

– ***Prior Information Effect on the Informed Ensembles***

To explore the effect of prior information to predict the accuracy when used greedy ensemble selection method, we chose subject J04 randomly. Figure 4.16 shows the results of the accuracy when constructed the informed ensemble based on the High Diversity. We found the performance on ensemble models increased when the diversity is increasing. By using the MSE as the prior information to build informed ensemble the Figure 4.17, we observed the accuracy increase when decreasing the result of the MSE. The Crowd error relies on the difference between the MSE and the diversity if we have low MSE or High Diversity that will reflect on the Crowd Error, the Figure 4.18 show the accuracy increased when the Crowd error decreased.

– ***The best type of Prior information to build informed ensembles***

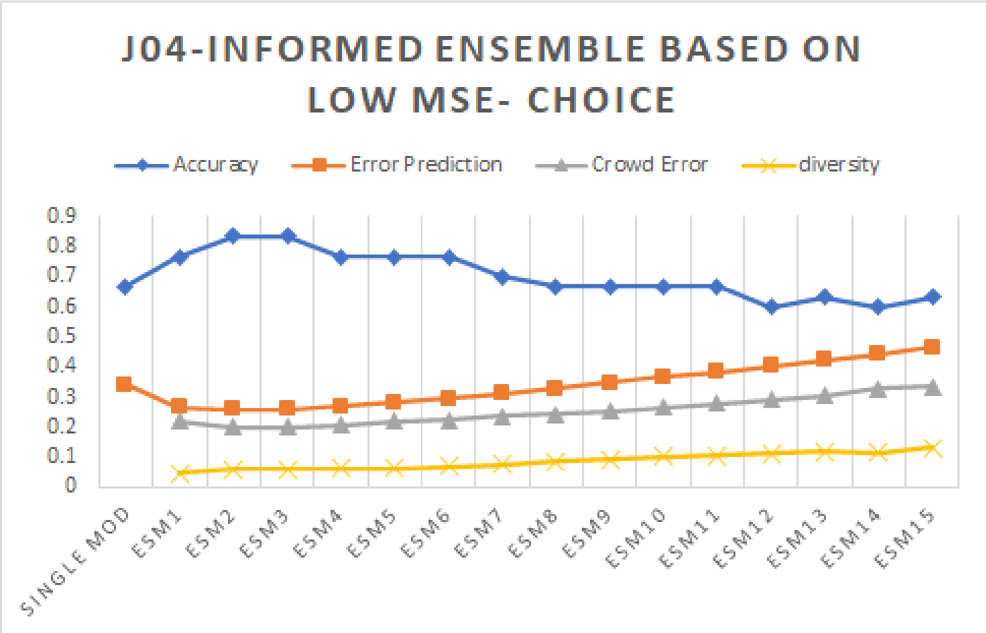


**Figure 4.17:** Informed Ensemble using Greedy Ensemble selection algorithm Based on MSE for the subject J04. ESM is mean the Ensemble Models.

– ***Accuracy performance*** For informed choice-based ensembles, three types of prior information used to build informed ensembles, high diversity, low MSE, and low crowd error were compared in terms of the accuracy of prediction. As a result, we chose the probability average (PA) as the aggregation techniques. The Greedy Ensemble Selection method explained in section 4.5.2 that used to build choice-based ensembles. Table 4.29 shows the results for common and personalized strategies to build informed choice-based ensembles based on the above prior information that is mentioned. Personalized strategy achieved the best performance for all prior information except high diversity, and the informed choice-based ensembles based on the Low Crowd Error present a high performance in both strategies.

– ***Frequency-Best-Model performance***

The term $f_{BestModel}$ provides the frequency of the rank of the best models among the type of prior information in both common and personalized strategies. We

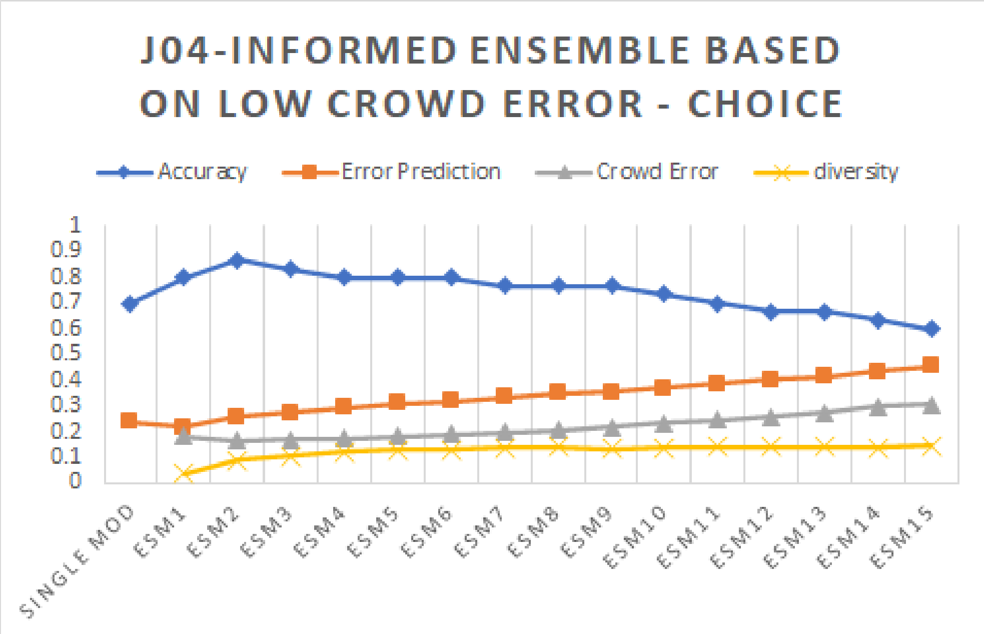


**Figure 4.18:** Informed Ensemble using Greedy Ensemble selection algorithm Based on Low Crowd Error for the subject J04. ESM is mean the Ensemble Models.

**Table 4.29:** Informed ensembles: comparison of common and personalized strategies. The PA aggregation technique was used in all cases.Three types of prior information used: High diversity, low MSE, and low crowd error. The average accuracy for informed choice-based ensemble.

| **Prior Information** | Strategy: **Personalized** | Strategy: **Common** |
|---|---|---|
| High Diversity | **0.71**(ns) | 0.64 |
| Low Crowd Error | **0.80**(ns) | 0.79 |
| Low MSE | **0.78**(ns) | 0.77 |

used it to do a comparison of these strategies. The rank of a common strategy better than the personalized strategy. and the informed choice-based ensembles based on the low MSE and low crowd error achieve to high performance which seems confirm to results when using the accuracy metrics (see Figure 4.19).

– ***Comparison informed ensembles with single learner***

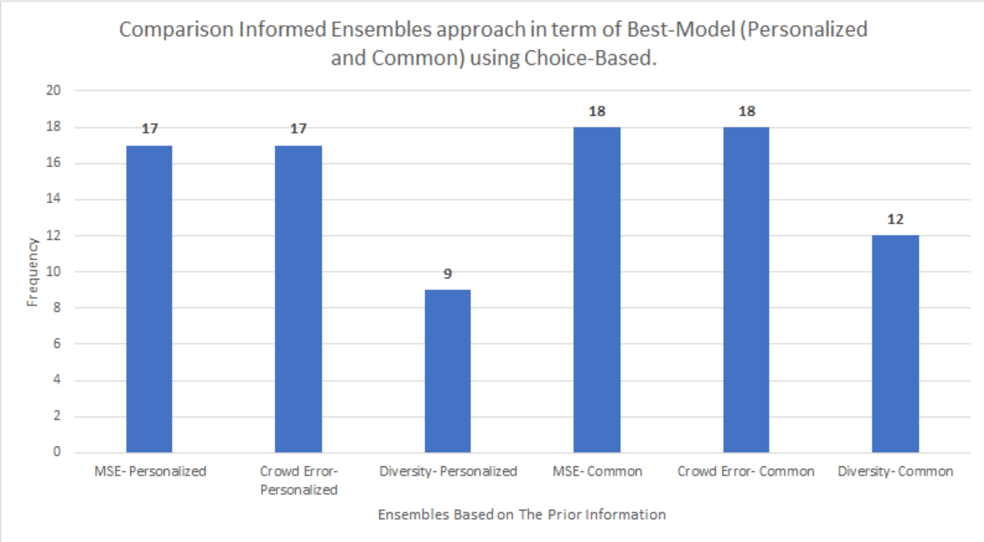


**Figure 4.19:** Comparison Informed Ensembles approach in term of Best-Model (Personalized and Common) using Choice-Based in term $f_{BestModel}$.

**Table 4.30:** Informed ensembles: comparison of choice-based ensemble with single choice models. The average accuracy over all users is shown.

| Prior Information | Informed Ensembles: Common Strategy | Single Learner |
|---|---|---|
| High Diversity | **0.64**(*) | 0.55 |
| Low MSE | **0.77**(***) | 0.62 |
| Low Crowd Error | **0.79**(***) | 0.63 |

– ***Accuracy performance***

Three types of prior information used to construct informed choice-based ensembles by greedy ensemble selection, High diversity, low MSE, and low crowd error, and common strategy used to apply the same model for all users, as well as the probability average (PA) as aggregation technique. Table 4.30 illustrates their performance when compared with their corresponding common best single model, the results show the performance on informed choice-based ensembles better single learner.

– ***Frequency-Best-Model performance***

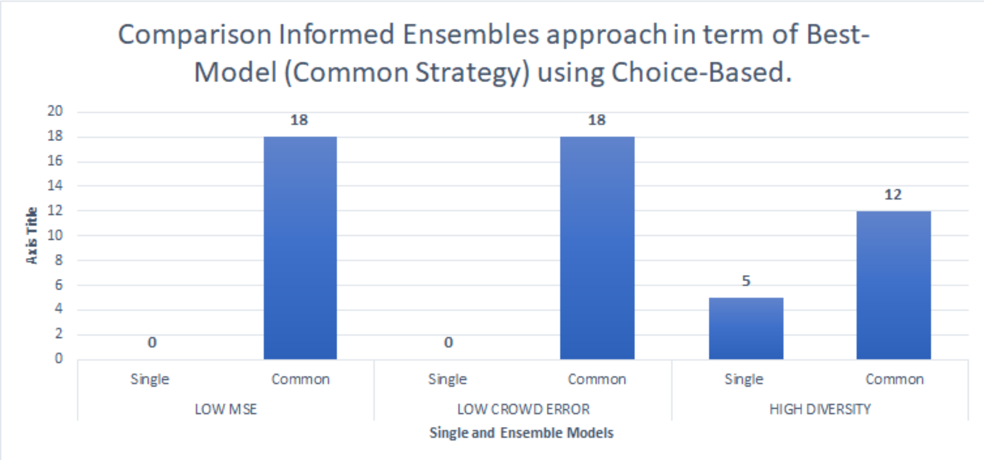


**Figure 4.20:** Comparison Informed Ensembles approach in term of Best-Model (Common Strategy) using Choice-Based.

The comparison between informed choice-based ensembles and single models in terms of $f_{BestModel}$ is in agreement with what it was found by using the accuracy metrics. Informed choice-based ensemble built using three different types prior information outperform their corresponding single choice-based model for all cases. And choice-based ensembles are the best for all subjects in both ensembles models built by the low MSE and crowd error ( see Figure 4.20).

– ***Comparison with ensembles built with Decision Trees.***

  – ***Accuracy performance***

    Three informed choice-based ensembles were built using three types of prior information, high diversity, low MSE, and low crowd error. All implemented the probability average (PA) as the aggregation technique. They compared with ensembles built with decision trees as a single learner. Table 4.31 shows the superior performance of informed choice based compared with DT for all cases, and the results show outperform the performance of the personalized strategy.

  – ***Frequency-Best-Model performance***

**Table 4.31:** Comparison between informed ensembles built with both common and personalized strategies using the prior information, high diversity, low MSE, and low crowd error. The performance of choice-based and DT-based ensembles in terms of average accuracy over all users is shown.

| | **Common Strategy** | | **Personalized Strategy** | |
|---|---|---|---|---|
| **Prior Information** | Learner: **Choice** | Learner: **DT** | Learner: **Choice** | Learner: **DT** |
| **High Diversity** | **0.64**(***) | 0.49 | **0.71**(***) | 0.53 |
| **Low MSE** | **0.77**(***) | 0.54 | **0.78**(***) | 0.55 |
| **Low Crowd Error** | **0.79**(***) | 0.52 | **0.80**(***) | 0.57 |

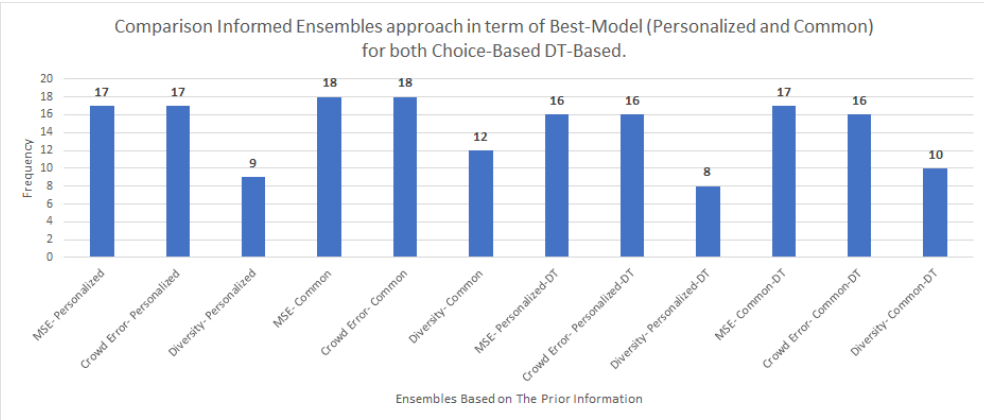


**Figure 4.21:** Comparison Informed Ensembles approach in term of $f_{BestModel}$, Personalized and Common Strategies for both Choice-Based DT-Based. Each column represents different experiment.

The comparison of choice-based and DT-based ensembles under both common and personalized strategies arrangement is shown in Figure 4.21 when analyzing the results. It's observed superior the rank of the choice-based models compared with the corresponding DT-models, that confirmed the accuracy performance metrics.

– ***Comparison between the Uninformed and Informed ensembles approaches***

**Table 4.32:** Comparison between informed ensembles with best 1-Learner and N-Learner ensembles by common strategy. The average accuracy over all users is shown.

| Prior Information | N-Learner: Personalized Top-2 | N-Learner: Common Top-2 | E-Choice-Bagging | R-Choice-Bagging | A-Choice-Bagging |
|---|---|---|---|---|---|
| High Diversity | **0.64(ns)** - 0.61 | **0.64(ns)** - 0.58 | **0.64**(***) - 0.38 | **0.64**(***) - 0.46 | **0.64**(*) - 0.55 |
| Low Crowd Error | **0.79**(***) - 0.61 | **0.79**(***) - 0.58 | **0.79**(***) - 0.38 | **0.79**(***) - 0.46 | **0.79**(***) - 0.55 |
| Low MSE | **0.77**(***) - 0.61 | **0.77**(***) - 0.58 | **0.77**(***) - 0.38 | **0.77**(***) - 0.46 | **0.77**(***) - 0.55 |

**Table 4.33:** Comparison between informed ensembles with best 1-Learner and N-Learner ensembles by personalized strategy. The average accuracy over all users is shown.

| Prior Information | N-Learner: Personalized Top-2 | N-Learner: Common Top-2 | E-Choice-Bagging | R-Choice-Bagging | A-Choice-Bagging |
|---|---|---|---|---|---|
| High Diversity | **0.71**(*) - 0.61 | **0.71**(**) - 0.58 | **0.71**(***) - 0.38 | **0.71**(***) - 0.46 | **0.71**(***) - 0.55 |
| Low Crowd Error | **0.80**(***) - 0.61 | **0.80**(***) - 0.58 | **0.80**(***) - 0.38 | **0.80**(***) - 0.46 | **0.80**(***) - 0.55 |
| Low MSE | **0.78**(***) - 0.61 | **0.78**(***) - 0.58 | **0.78**(***) - 0.38 | **0.78**(***) - 0.46 | **0.78**(***) - 0.55 |

– ***Accuracy performance***

Two strategies used to build the Blind (uninformed) ensemble, 1-learner, and N-learners ensembles. For 1-learner three choice-based ensembles were built using bagging sampling. Regard to N-learner we chose the best Top models: Top-2, and the comparison made by Common and Personalized strategies. With respect to an informed choice-based ensemble, three prior information used to construct the ensembles models: high diversity, low MSE, and low crowd Error, and two strategies were used, common and personalized. also, we chose the probability average (PA) as an aggregation technique. Table 4.32 shows outperform the performance of an informed choice-based ensemble compared with uninformed ensembles in all cases. The 1-Learner offers the worst performance and superior performance of the informed ensembles using personalized strategies, and the difference is significant in all cases( see Table 4.33).

– ***Frequency-Best-Model performance***

The comparison made by all the models that mention in the previous Accuracy performance section, and the evaluation of models by the term of $f_{BestModel}$. The comparison of the informed choice-based ensembles and uninformed choice-based ensembles shown in figure 4.22. When analyzing the results we observed outperform the informed ensembles on uninformed ensembles by using the low MSE

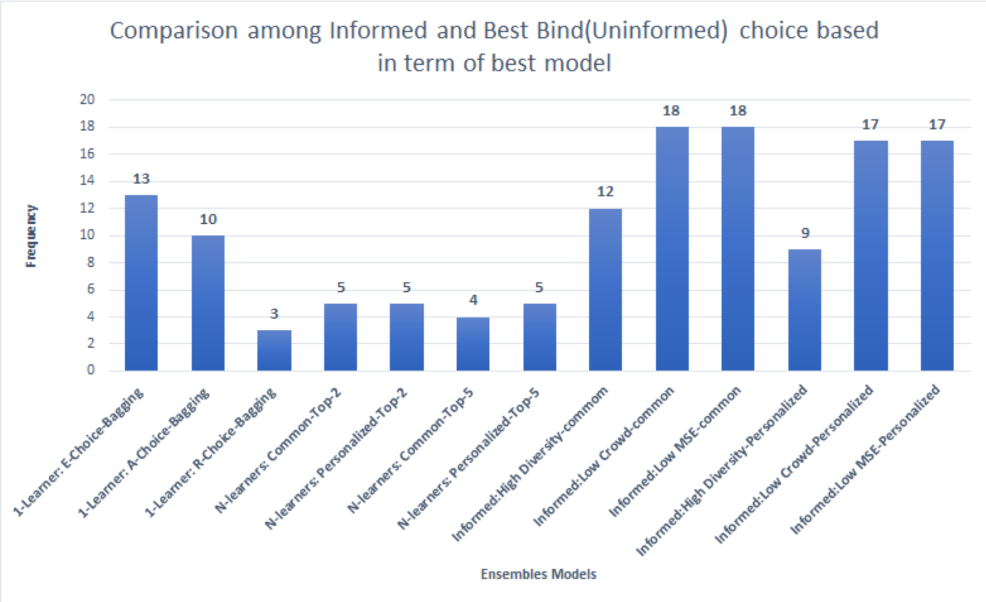


**Figure 4.22:** Comparison among Informed and Best Bind(Uninformed) choice based in term of $f_{BestModel}$,Common and Personalized strategies, and High diversity, Low MSE, and Low Crowd Error as prior information for informed method.Each column represents different experiment.Each column represents different experiment.

and crowd error as prior information with both common and personalized strategies. 1-Learner build using emotional choice bagging outperformed on informed ensemble when using the high diversity as prior information. we observed the worst result when building the N-Learner models.

# Chapter 5

# Discussion

With the aim of providing transparency and interpretability, we have developed choice-based models, which use sound decision-making principles to guide the recommendation process. From this perspective, the recommendation problem is considered a problem of choice prediction rather than rating prediction, the current paradigm in the field of recommender systems. The key elements of choice-based models are as follows: (1) user's preferences are learned from choices, (2) the choice set of each choice situation is treated as an important variable for both explaining and predicting future choices, and (3) unobserved factors affecting the decision-making process are captured through random variables. Recent content-based approaches [8] share the same idea about the potential of using choices to derive preferences, but are limited to pairwise comparisons rather than comparisons of full choice sets.

The approach proposed in this thesis poses the question about the best source of data to generate recommendations: ratings or choices (Q1, H1). Two criteria have been used to answer this question: accuracy of learned preferences, and interpretation of predictions. On the one hand, the use of choices and choice-based modelling may solve the issue of accuracy of the preferences learned from ratings. While ratings represent a post-experience outcome that may be more dependent on the satisfaction of previous users expectations rather than on preferences, choices are the result of the direct matching between the user's preferences and the item's attributes. Therefore, choice data may be an accurate source from which preferences might be learned. On the other hand, choice models also overcome the interpretability issue, as the estimated model coefficients provide a means of easily explaining why some items are more likely to be recommended than others.

The next step was to compare the performance of choice-based models with rating-based algorithms and basic ensembles (Q2, H2). The benefits of the choice-based approach in terms of preference's accuracy and interpretability do not dramatically affect performance. The results shown in tables 4.5, 4.6, and 4.7, and summarized in Figure 4.8, suggest that choice models may be inferior to ensembles only in situations where larger data sets are available. The results confirm our hypothesis H2 based on the frame of the data (fewer variables users and less alternatives the choice set). The drawback of gathering information about the domain (an item's attributes and the values of these) is compensated for in two ways: (1) by using more accurate data as choices represent real actions, while ratings are subjective values that are very sensitive to mood and context states, and (2) by removing the burden of interrogating decision-makers about their post-experience satisfaction.

At this point, it is important to understand that the problem of prediction of human choices is naturally related to the problem of human decision-making and cognition. The pieces of evidence from fields like behavioral economics, neuroscience, and neuromarketing suggest that System I circuits driven by emotional factors are the ones in control of most of our daily choices. The motivation of our study was to test this line of thinking by developing a choice-based models enriched with relevant cognitive variables that could explain the observed behavior.

With regard to the best cognitive choice-based model (Q3, H3), the results are shown in Figure 4.9 to confirm that the first part of H3, i.e., both S-I-Attentional and S-I-Emotional models provide superior performance in terms of the frequency of the best model. However, when the average accuracy for all users is considered as metrics (see Table 4.8), the S-II-rational model behaves better than all the S-I-Emotional models. In fact, the second part of H3 was found to be incorrect. The S-I-Attentional models in general, and the A2 model in particular, they made a better job than all S-I-Emotional models. These findings suggest that the factors underlying the attention process are playing a relevant role in the decision-making process.

As for the quest of the best cognitive choice-based model for all users (Q4, H4), the accuracy results are shown in Table 4.9 seem to contradict H4. The best model on average

(A2 model) is not the best model for all users. Not only that, but it may also perform badly for a significant number of subjects. Furthermore, the data indicate a great variety of best models, including all different types of attentional, emotional, and rational ones. These evidences seem to point out the richness of cognitive diversity and the existence of a set of different mindsets that could take control of human behavior under different contexts.

The issue of the performance of ensembles (Q5, H5) has been faced with two different approaches. In the first one, we built ensembles with separated subsets of features. Under this scheme, the results shown in section 4.4.3 from Table 4.10 to Table 4.15 indicate those ensembles cannot overcome the performance of single choice models. So, H5 is wrong under the first approach. In the second one, we built ensembles by aggregating all the features. In this case, the results shown in Tables 4.17 and 4.18 partially support H5. The combination of all features seems to allow ensembles to outperform the single choice models. This finding may point out that both S-I and S-II mechanisms could operate together as an ensemble to determine human choices.

With regard to the best ensemble for all users (Q6, H6), the evidences shown in Table 4.16 clearly contradicts H6. There is no ensemble being the best for all subjects, which seems coherent with the answer found for Q4.

The line of work went further with the hybridization of high-performance ensembles with transparent and interpretable choice-based models. Specifically, we have wondered about the optimal methods to build choice-based ensembles. Two strategies were analyzed to select the set of learners in the ensemble in the most efficient way: uninformed, or blind, as well as informed methods. With regard to an uninformed method, Several factors to build the ensembles have been explored. It included different dimensions of the problem: the number of different learners in the ensemble, the sampling strategy, the aggregation technique, and the degree of personalization.

On testing the best method to build uninformed choice-based ensembles (Q7, H7), we have found evidences that confirm hypothesis H7. The N-Learners ensembles provided the best average performance as it was shown in Table 4.22. Among them, the personalized strategy was superior to the common strategy, which makes sense if we recall that in the former the learners of the ensemble are specifically chosen for each subject. However, when turning the point of view using the $f_{BestModel}$, the 1-Learner ensembles, surprisingly, make a better job than N-learners ensembles. A possible explanation to reconcile the results of the two metrics is that 1-Learner ensembles were providing high accurate predictions on a set of subjects but poor ones on the other ones. On the other side, N-Learners ensembles may seem to generate more consistent predictions over the full set of subjects. Another important finding was the observation that the diversity provided by adding weaker learners to the ensemble was not able to balance the loss of average accuracy resulted in that addiction (see Table 4.21).

With regard to the performance of uninformed choice-based ensembles compared to single choice models (Q8, H8), hypothesis H8 was also confirmed. Both 1-Learner and N-Learners ensembles generate better predictions than the single choice-based models in terms of accuracy. Under the $f_{BestModel}$ metrics, the pattern is not so clear. 1-Learner models are clearly better using the bagging sampling strategy, but are outperformed by single models when using the boosting approach. This might indicate that boosting is not an efficient method to build choice-based methods. On the other hand, N-Learner models do a better job than single ones but only in the Top-2 arrangement. The aggregation of weaker learners to the ensembles (Top-5 to All) reduces the performance of the ensembles in a significant way.

As for the performance of informed choice-based ensembles compared to ensembles built with other models (Q9, H9), the obtained results support H9. Choice-based ensembles outperform DT-based ones in the majority of configurations for both 1-Learner and N-Learners arrangements. This rule has an exception when using the boosting sampling strategy for 1-Learner ensembles. The effect of boosting on pushing the performance of DT-based ensembles might indicate that DT learners are weaker and more easily improved than choice learners.

The work aimed to examine the best prior information type in order to build informed choice-based ensembles. It included different types of prior information: high diversity, low error prediction (MSE), and low crowd error.

With regard to the best method to build informed choice-based ensembles (Q10, H10), we found a contradiction with a hypothesis H10. The best results were generated with informed ensembles using Low Crowd Error, and the worst way by using high diversity as prior information as it was shown in Table 4.29. Also, those results confirmed when changing the point of view using the $f_{BestModel}$. The identical results appeared by applying common and personalized strategies.

Concerning the performance of informed choice-based ensembles compared to single choice models (Q11, H11), Hypothesis H11 was confirmed. Informed choice-based ensembles generate better prediction than the single choice-based models. That is obvious when using the term the accuracy performance for all informed ensembles as it was presented in Table 4.30. Moreover, in the term $f_{BestModel}$, the ensembles models are the best for all subjects when built informed ensemble by low crowd error ( see Figure 4.20). As for the performance of informed choice-based ensembles compared to ensembles built with other models (Q12, H12), the obtained results verified the hypothesis H12. Informed choice-based ensembles outperform the DT-based in both common and personalized strategies under two metrics, which are used the accuracy performance and $f_{BestModel}$ for all corresponding cases as it were shown in Table 4.31 and Figure 4.21, which seems coherent with the answer found for Q9.

Regarding with the best method to build optimal choice-based ensembles (Q13, H13), Hypothesis H13 was also confirmed. The results in the performance of informed choice-based outperform on uninformed choice-based ensembles in both terms of accuracy and $f_{BestModel}$ for all ensembles choice-based models as it were shown in Tables 4.32, 4.33 and Figure 4.22.

# CHAPTER 6

# CONCLUSIONS

The work aimed to present new techniques and methods to develop NEW TECHNIQUES AND METHODS TO DEVELOP RECOMMENDER SYSTEM ALGORITHMS based on the use of the choice-based models and applied choice-based ensembles to provide more accuracy and explained the decision-maker behavior. In this thesis, we have presented a set of publications showing how to develop choice-based models and apply them to Recommender Systems. We consider that we have adequately succeeded in this goal by achieving different conclusions.

By using the choice-based model approach, the limitation of the accuracy of the preferences learned from ratings is overcome. Furthermore, the new models seem to provide an optimal trade-off between interpretability and performance, which paves the way to the application of more complex decision-making models in the field of recommender systems. On the way to build better models, we resorted on decision-making theories as well as cognitive drivers of behaviour. We have discovered interesting features by building cognitive choice-based models: (1) S-I-Emotional models were not the best predictors of choice behavior, and (2) there is no such thing as the best choice-based model or the best choice-based ensembles for all users.

We also wondered about the optimal way to build choice-based ensembles by using single cognitive models. On one side, uninformed methods provided good results and illustrated the importance of both sample strategies and aggregation techniques. Two main strategies, 1-Learner and N-Learners type ensembles were analyzed in terms of their accuracy as well as the frequency of best-model metrics. The results pointed out the superior performance of N-Learners ensembles and showed the potential of personalized arrangements. On the other side, informed methods were analyzed and tested in order to optimize the selection process of each new learner in the ensemble. The new informed ensemble models, fueled by prior information, specifically the crowd error, showed very promising results and outperformed uninformed choice-based ensembles. We believe that these findings will motivate not only the development of a new area of recommendation algorithms, but also an exciting pathway to explore the interplay and aggregation of cognitive modules in human brains.

The research carried out in the doctoral thesis has just opened a new line of work in the fields of Recommender Systems and Decision-making. To improve and generalize the results, our future work will focus on ecological experiments to include other drivers that may explain human behaviour. The list of candidates include: personality traits, knowledge level, personal state, and context factors. The problem of predicting human behaviour is complex, but we are on the way to understand it better and to apply this knowledge to facilitate decisions and choices in the digital world.

# Appendix A

# Appendix

## A.1 Rectur Voting Form

The Santiago(é)Tapas contest included many restaurants and the tapas offered by these restaurants. During the contest, each participating restaurant had coaster-shaped voting forms ( see Figure A.1 ). After tasting a tapa, the tourist filled in the voting form and put it in a specific box. This voting form asked about the user's demographics, asked about the evaluation of the tapa and restaurant using a Likert scale, and gathered info about the context of the experience, specifically with whom and when the experience occurred. This information was entered into the database.

The coaster voting form evaluation data obtained after the contest. The evaluation profile data included information on user ID, DNI, tapa number, tapa rating (evaluation of tapa using the Likert method, where 0 does not like at all, and 5 is like a lot), service rating, restaurant value, global experience rating, and restaurant ID.

## A.2 EEG Recording with EMOTIV EPOC

The signal recorded on the EEG comes from the brain's electrical activity as a result of nerve impulses (electrical currents) generated by neurons in the cerebral cortex. This collected electrical signal is amplified and represented in the form of lines, interpreting the activity of the different brain areas over time.


**Figure A.1:** Coaster voting form for Santiago(é)Tapas contest .

The helmet that was used for the acquisition that the activity was Emotiv COPD. It consists of 16 gold-plated stainless steel electrodes, hydrated with a saline solution to lower the impedances to the level required to improve contact with the skin's surface. Fourteen electrodes were placed at positions AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4 and two reference CMS and DRL at locations P3 / P4, according with the International System 10-20. Figure A.2 shows an example of the configuration of the electrodes and the used helmet Emotiv EPOC.

The Emotiv EPOC helmet connects with the computer through a Bluetooth connection with a Lithium battery with a range of 12 hours. It has a signal pickup frequency of 128Hz. A 2-30 Hz bandpass filter and a 50Hz Notch filter were applied, Figure A.3.

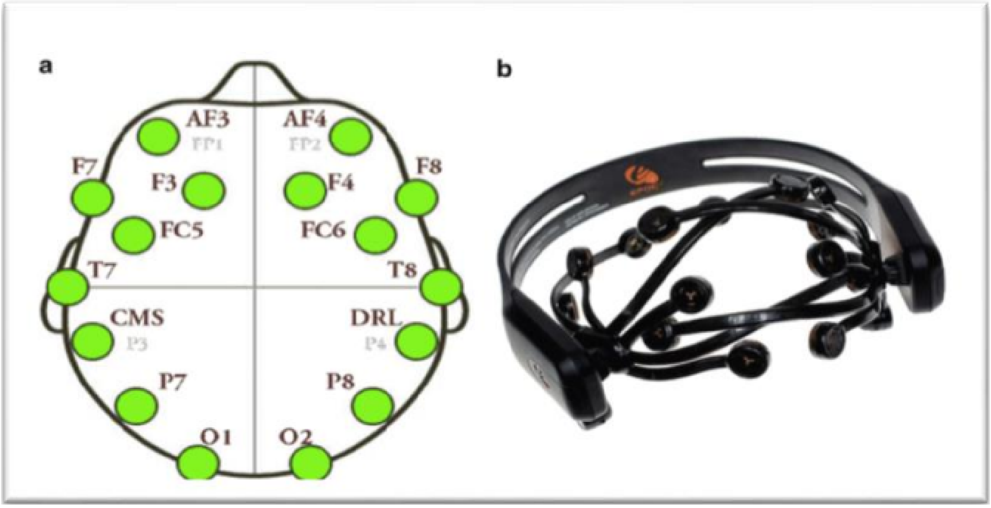


**Figure A.2:** a) Distribution of the electrodes according to the System International 10-20. b) Example of an Emotiv EPOC helmet.

The EEG data were subsequently separated into four frequency bands: Theta (4-8 Hz), Alpha (8-12 Hz), Beta (12-25 Hz), and Gamma (> 25 Hz), which are used to examine cognitive or affective processes in response to the choices shown. The Emotive EPOC devices allow you to collect the following metrics:

- Engagement/Boredom reflects long-term alertness and the conscious direction of attention towards task-relevant stimuli.
- Excitement (Arousal) reflects the instantaneous physiological arousal towards stimuli associated with positive valence.
- Stress (Frustration).
- Meditation (Relaxation).


**Figure A.3:** Emotiv EPOC helmet connects with the computer.

## A.3 Facial coding with FACET

While EEG alone already provides an incredibly rich amount of insights into cognitive, affective, and attentional foundations of human behavior, you might want to consider adding other sensors to get the entire picture. Therefore, the addition of facial tracking was considered.

The FACET captures facial movements using a high-resolution camera capable of recognizing the slightest gesture related to emotions. It is not intrusive; evaluates the position and orientation of the head (ideal for the identification of artifacts of the rocking and head movements); micro-expressions (such as raising the eyebrows or opening the mouth) and global facial expressions of basic emotions (see Figure A.4). The camera is placed in front of the participant and on top of the screen.

The camera that was used was Logitech HD Pro Webcam C920, which has high quality 1080p HD video capture (up to 1920 x 1080 pixels) and low light auto-correction.

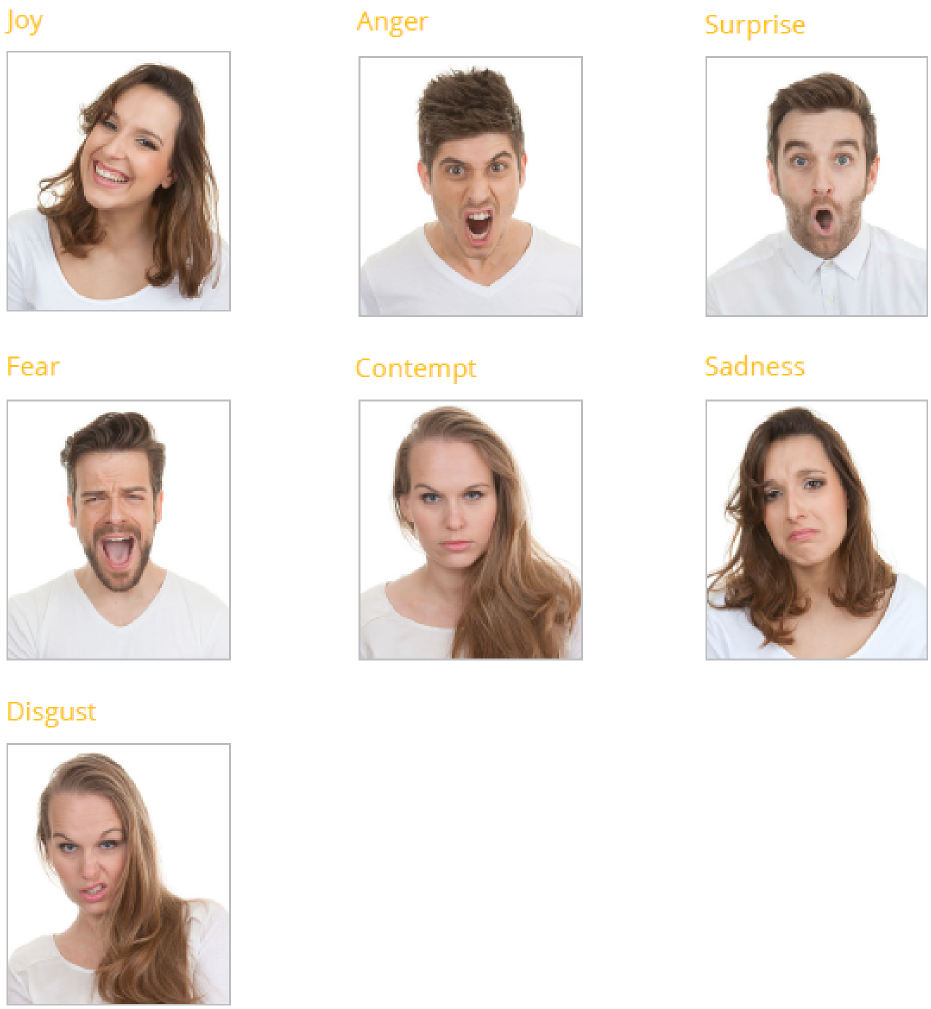


**Figure A.4:** Global facial expressions of basic emotions.

## A.4 Eye-tracking with TOBII

An eye-tracking device was required for eye tracking, detects and records eye position and pupillary dilation during the tasks of the experiment, Figure A.5. The Tobii X2-30 device operating at 30Hz (30 photos per second of each eye) was used. It was placed under the

**Figure A.5:** Head mounted eye trackers.

screen to detect the gaze situation with the previous calibration in each participant, which consisted of following the red dot on the screen with the gaze.

Eye-tracking was used to analyze the sequence of the patterns, which gave us a more in-depth insight into the cognitive processes underlying care. The most common metrics used in eye tracking research and what you can make of them:

- Fixation and Gaze Points: Gaze points are the basic unit of measurement - a gaze point equals a raw sample captured by the eye tracker. If a series of gaze points becomes close in time and range, the resulting gaze group denotes a fixation, a period in which our eyes are locked onto a specific object. Typically, the duration of fixation is 100 to 300 milliseconds, which is the time it takes for the brain to process information,Figure A.6.
- The eye movements between the fixations are known as "saccades" which are very fast eye movements.


**Figure A.6:** Fixation and Gaze Points.

– Time spent: Time spent quantifies the amount of time that respondents have spent on an Areas of Interest (AOI).

Visual information is generally only perceived during fixations, not during exits. Fixings and saccades are excellent measures of visual attention and interest.

# Bibliography


[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

[2] Charu C Aggarwal et al. *Recommender systems*, volume 1. Springer, 2016.

[3] Ameed Almomani, Cristina Monreal, Jorge Sieira, Juan Graña, and Eduardo Sánchez. Rational, emotional, and attentional choice models for recommender systems. In *World Conference on Information Systems and Technologies*, pages 557–566. Springer, 2019.

[4] Ameed Almomani, Paula Saavedra, and Eduardo Sánchez. Ensembles of decision trees for recommending touristic items. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 510–519. Springer, 2017.

[5] Ameed Almomani and Eduardo Sánchez. Uninformed methods to build optimal choice-based ensembles. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*, pages 58–65. Springer, 2019.

[6] Marko Balabanović and Yoav Shoham. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66–72, 1997.

[7] Moshe E Ben-Akiva, Steven R Lerman, and Steven R Lerman. *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, 1985.

[8] Laura Blédaité and Francesco Ricci. Pairwise preferences elicitation and exploitation for conversational collaborative filtering. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 231–236. ACM, 2015.

[9] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[10] L Breiman, JH Friedman, RA Olshen, and CJ Stone. Classification and regression trees. 1984. wadsworth. *Inc., Monterey, Calif., USA*.

[11] Robin Burke, Alexander Felfernig, and Mehmet H Göker. Recommender systems: An overview. *Ai Magazine*, 32(3):13–18, 2011.

[12] Yves Croissant et al. Estimation of multinomial logit models in r: The mlogit packages. *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf*, 2012.

[13] Antonio R Damasio. The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1346):1413–1420, 1996.

[14] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.

[15] Mehdi Elahi, Matthias Braunhofer, Francesco Ricci, and Marko Tkalcic. Personality-based active learning for collaborative filtering recommender systems. In *Congress of the Italian Association for Artificial Intelligence*, pages 360–371. Springer, 2013.

[16] Banda Gerald. A brief review of independent, dependent and one sample t-test. *International Journal of Applied Mathematics and Theoretical Physics*, 4(2):50–54, 2018.

[17] David Goldberg, David Nichols, Brian M Oki, and Douglas Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.

[18] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning: data mining, inference, and prediction, springer series in statistics, 2009.

[19] Jonathan L. Herlocker and Joseph A. Konstan. Content-independent task-focused recommendation. *IEEE Internet Computing*, 5(6):40–47, 2001.

[20] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 241–250, 2000.

[21] Rong Hu and Pearl Pu. A study on user perception of personality-based recommender systems. In *International conference on user modeling, adaptation, and personalization*, pages 291–302. Springer, 2010.

[22] Jasur Ismoilov. *Stated and revealed preferences on gastronomic tourism in Santiago de Compostela: PHD Thesis*. PhD thesis, Universidade de Santiago de Compostela, 2017.

[23] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[24] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[25] Daniel Kahneman. *Thinking, fast and slow*. Macmillan, 2011.

[26] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

[27] Yehuda Koren. The bellkor solution to the netflix grand prize. *Netflix prize documentation*, 81(2009):1–10, 2009.

[28] Wee Sun Lee. Collaborative learning for recommender systems. In *ICML*, volume 1, pages 314–321. Citeseer, 2001.

[29] Martin Lindstrom. *Buyology: How everything we believe about why we buy is wrong*. Random House, 2012.

[30] Linyuan Lü, Matúš Medo, Chi Ho Yeung, Yi-Cheng Zhang, Zi-Ke Zhang, and Tao Zhou. Recommender systems. *Physics reports*, 519(1):1–49, 2012.

[31] Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

[32] Scott E Page. Where diversity comes from and why it matters? *European Journal of Social Psychology*, 44(4):267–279, 2014.

[33] Fábio AP Paiva, José AF Costa, and Cláudio RM Silva. A personality-based recommender system for semantic searches in vehicles sales portals. In *International Conference on Hybrid Artificial Intelligence Systems*, pages 600–612. Springer, 2017.

[34] Robi Polikar. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45, 2006.

[35] Anantha Krishnan Pradeep. *The buying brain: Secrets for selling to the subconscious mind*. John Wiley & Sons, 2010.

[36] Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: introduction and challenges. In *Recommender systems handbook*, pages 1–34. Springer, 2015.

[37] Paula Saavedra, Pablo Barreiro, Roi Duran, Rosa Crujeiras, María Loureiro, and Eduardo Sánchez Vila. Choice-based recommender systems. In *RecTour@ RecSys*, pages 38–46, 2016.

[38] J Ben Schafer, Joseph Konstan, and John Riedl. Recommender systems in e-commerce. In *Proceedings of the 1st ACM conference on Electronic commerce*, pages 158–166, 1999.

[39] John Scott. Rational choice theory. *Understanding contemporary society: Theories of the present*, 129:671–85, 2000.

[40] Amartya Sen. Rational behaviour. In *Utility and probability*, pages 198–216. Springer, 1990.

[41] Geetha Sitaraman. *Inferring big 5 personality from online social networks*. PhD thesis, 2014.

[42] Marko Tkalcic, Matevz Kunaver, Jurij Tasic, and Andrej Košir. Personality based user similarity measure for a collaborative recommender system. In *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real world challenges*, pages 30–37, 2009.

[43] Kenneth E Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

[44] Grigorios Tsoumakas, Ioannis Partalas, and Ioannis Vlahavas. A taxonomy and short review of ensemble selection. In *Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications*, pages 1–6, 2008.

[45] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

[46] Stephen Vaisey. Motivation and justification: A dual-process model of culture in action. *American journal of sociology*, 114(6):1675–1715, 2009.

# List of Figures

# List of Tables

# LIST OF ABBREVIATIONS

**RSs** Recommender Systems

**DM** Decision Making

**S-I** System I

**S-II** System II

**CF** Collaborative Filtering

**DT** Decision Trees

**CBF** Content-based Filtering

**SVD** Singular Value Decomposition

**RUMs** Random Utility Models

**RF** Random forests

**CITIUS** Research Center on Information Technologiesat USC

**EEG** Electroencephalography

**FACET** Facial Coding

**R** Rational Model

**A1** Attentional Model 1

**A2** Attentional Model 2

**E1** Emotional Model 1

**E2** Emotional Model 2

**E3** Emotional Model 3

**E4** Emotional Model 4

**E5** Emotional Model 5

**E6** Emotional Model 6

**E7** Emotional Model 7

**E8** Emotional Model 8

**E9** Emotional Model 9

**E10** Emotional Model 10

**E11** Emotional Model 11

**E12** Emotional Model 12

**E13** Emotional Model 13

**PA** Probability Average

**CF-UB** Collaborative Filtering- User-based

**CF-MF** Collaborative Filtering-Matrix Factorization

**CART** Classification and Regression Tree

**RMSE** Root Mean Squared Error

**DCG** Discounted Cummulative Gain

**MSE** Discounted Cummulative Gain

**CSE** Crowd Square Error

**R.CV** Random cross validation

**LOO.CV** leave-one-out cross validation

**J02, V01** labels indicating day and experimental slot

**R.CV** Gradient boosting models

**DT** Decision Tree