Convocatoria 2020 - «Proyectos de I+D+i»

IP 1 (Nombre y apellidos)**:** Lara Quijano Sánchez

IP 2 (Nombre y apellidos)**:** Iván Cantador Gutiérrez

TÍTULO DEL PROYECTO (ACRÓNIMO): Librería de Potenciamiento de Tweets (STE)
TITLE OF THE PROJECT (ACRONYM): Suite for Tweet Enhancement (STE)

1. Propuesta científica – scientific proposal

1.1 Background, current status and justification of the proposal

1.1.1. Background

Twitter is one of the most common micro-blogging sites on the Internet, with over 187 million active users posting short messages or *tweets* every day (Tankovska [2021]). The character limit of a tweet (280 characters) encourages users to casually upload new content, whereas conventional blogging requires more time investment to write a new message. Furthermore, since more and more people own mobile devices around the world, Twitter events are becoming increasingly big, resulting in over 500 million tweets being posted to the social network each day. The downside of Twitter's success is that users can become quickly overwhelmed by the large amount of data to which they are exposed on a daily basis (Otsuka et al. [2016]). As a mechanism to combat this issue, Twitter users have assimilated the usage of hashtags to filter which content they are exposed to. A *hashtag* is a word or expression that is preceded by the hash symbol "#" and can be used anywhere in the body of a tweet. Tagging tweets with hashtags allows trends to spread rapidly among millions of users, forming an immediate group of people with common interests. Since the introduction of Twitter's hashtag search tool, many individual users and company marketers have begun using tagging to group tweets into similar discussions in order to facilitate searching for posts using the associated hashtags. But the "#" symbol is not the only relevant one on social networks, and more specifically on twitter. The "@" symbol can be used with the tweet to mention users whom the user wants to alert regarding the message. By mentioning users in a post, they will receive notifications and their possible retweets may help to initiate large cascade diffusion of the tweet. To enhance a tweet's diffusion by finding the right persons to mention and improve communication efficiency an automatic suggestion to the user of a small list of candidate names can be used (Gui et al. [2019], Wang et al. [2013]). Apart from the mentions and hashtags, the tweet content is a critical point while posting: the way it has been written may make that the impressions levels rise above unsuspected limits or by the opposite the tweet goes unnoticed (Zhao and Lan [2015]).

One of the most attractive features of the microblogging sites is the fact that events of widespread interest to a group of users result in a surge in real-time mentions as they occur. For example, users live-tweet about sport events, discuss about breaking news, or celebrate certain events on a memorial day, among others. Consequently, there is a spike in behavior correlated with the event in question, resulting in a social trend. While these social trends can reveal what is going on at the very moment and help discover certain events early on, a list of trends includes just the set of terms that are being mentioned outstandingly at the moment. Therefore, the time a tweet is posted is critical for the success or not of the publication. A prediction of the most suitable moment to publish our messages can improve the impact. One of the best tools to join current trends and achieve more popularity is the hashtag: similar tweets can be found by using it, and also more users will be able to see publications. In some social networks such as Twitter or Instagram, it is possible to follow hashtags or topics: although there is no follow-up relationship with other users, they will see the content tagged with the hashtags followed, so that if the appropriate terms are used, a high degree of popularity can be achieved. The importance of the hashtags in the tweets' popularity has been analyzed and established in many research papers, e.g. Lahuerta-Otero and Cordero-Gutiérrez, Pancer, Kostygina et al. [2020], Oliveira. Another mechanism to reach as many people as possible and join trends is to make the most appropriate mentions according to the tweet content. Since 2019 Twitter and other microblogging sites offer options such as "Related to a user/topic/hashtag", so, if the appropriate mentions are made, as with hashtags, it will be possible to reach a large number of users with which there is not even a followee relationship. The importance of the mentions has been demonstrated in Gui et al. [2019], Wang et al. [2013].

In this new social atmosphere, the trending phenomenon has arisen: some hashtags gain traction rapidly while other similar hashtags remain unseen and become irrelevant. Controlling and predicting this behavior has become vital in many businesses (Schaupp and Bélanger [2014], Pérez-Serrano et al. [2020], Dormanesh et al.[2020]), political campaigns Owen [2017] or discussions in general. There are even hundreds of jobs are being generated as a community manager: people who manage social networks accounts and try to get better opinions from users, promote products, influence voting decisions, etc. In this work, a new system for automatically recommending customized trending tweets based on users' input is proposed, with the objective of helping those who intend to boost their popularity (regular users, community managers, politicians, etc.) to increase the reach of their tweets. The proposed system will be made up of several independent modules: hashtag, user tagging, tweet paraphrasing or time of publication recommendation and popularity prediction for each of the recommendations made, which together will constitute the Suite for Tweet Enhancement: STE.

On the other hand, Machine learning (ML) (Murphy, [2012]) is present in many technologies of modern society. And society is enjoying this technology in the form of much better products and services. One of the key developments in this new golden era of Artificial Intelligence (AI) is deep learning. Deep learning (Goodfellow et al., [2016]) is behind most of the recent success of the field in the last years. In the proposed project these novel technologies will be employed. By this way, the recommendations and predictions made will have very high levels of success obtaining a suite that we expect it to become a standard in any area: business, politics, leisure, etc.

Differently from existing similar frameworks, our system allows the visualization of recommendations ordered according to the popularity prediction made. This paper's presented framework not only chooses and/or generates the improved tweets and returns them to the user, but the text strings obtained are ordered with the help of a popularity prediction module and are returned in descending order of predicted popularity. There are some works which rank different tweets or users according to their popularity using different metrics Aleidi et al. [2019], Yadav et al. [2017], Iqbal et al. [2021], Bacha and Thi Zin [2018], but all of them do it with already published tweets. In this way, the user will be able to choose hashtags not only based on the text displayed but also on the impact that the new tweet could potentially have. Another important aspect is that all this is achieved by simply entering the text in a text box and clicking the Get Recommendation button, making it simple and straightforward. The importance of providing a clean and user-friendly interface to our system is critical, as analyzed for example in Ali et al. [2015], Jitnupong and Jirachiefpattana [2018]. In this way, any kind of stakeholders (i.e individual users, companies or governments) will be able to improve their publications without any effort or difficulty. The expected user workflow is shown schematically in the next figure.

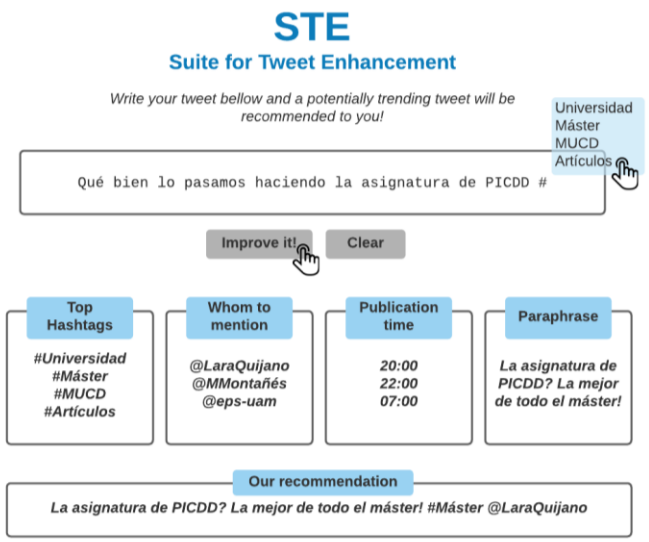

Figure 1.1.1. System usage example.

As a metric to measure a single tweet's popularity, we will use a mix of likes (l), retweets (r) and interactions (i), similar to Gao et al. [2021]. Specifically, the proposed formula for a tweet t is:

$$popularity(t) = r(t) + l(t) + i(t)$$

1.1.2. Current status

Twitter is one of the most common micro-blogging sites on the Internet. Thus, there is a lot of work around it, most of it related to classification techniques for opinion mining and sentiment analysis (Hemmatian and Sohrabi, 2019). A big part of them uses models based on LDA for topic modeling (around 27% e.g., Silva et al. [2019], Geetha and Karthika [2020], Osorio-Arjona et al. [2021]), which is the first building block in our system. Topic modeling concerns itself with the extraction of common topics from a collection of documents, or in our case, tweets. On the other hand, a surprisingly but encouraging realization was the small number of articles that delved into text generation in hashtag recommendation, mentions or publication time recommendation and paraphrasing tweets, only about 2% of them (Çetinkaya et al., [2020]). This proves that a deep look into these topics is certainly not superfluous. Our work is focused on this gap in research found, i.e., mentions, publication time and hashtag recommendation via automatic text generation combined with popularity prediction in Twitter. Also, a paraphrasing option will be included. The remaining of this section describes the current state of the art in different sub-areas inside our main area of research.

1.1.2.1. Predicting popular content

With the arrival and consolidation of social networks, the number and quality of research studies focused on them has risen accordingly, with several studies concentrating on the issue of predicting popular content. In this project the application domain considered is Twitter, where users and companies can greatly benefit from tools to predict whether their tweets are going to be popular and adjust the content accordingly.

Among the reviewed existing work in popularity prediction (without recommendation), relatively little has been done to study the relationship between topics and hashtags to predict tweets popularity. For example, in Roy et al., [2020], the authors proposed a Recurrent Neural Network based hybrid model with Feature Concatenation (RNN-HMFC), an approach that allowed them to predict the future popularity of online posts with an F1-score of 91.60% in the SEISMIC dataset. This is a dataset used in several popularity prediction research studies (e.g., Gou et al. [2018], Kefato et al. [2018]), which consists of approximately 3 billion tweets and a little over 160.000 tweets labeled as being popular. Also, a recent study on predicting a tweets popularity is Sun and Gloor, [2021], which forecasts the popularity for the next day using the number of mentions in the past week of a hashtag. They repeated their experiment in two different datasets consisting of 200k and 700k extracted from Twitter, and in a third article-based dataset consisting of 150k documents (Misra [2018]). They achieve a *Root Mean Square Error* of 10.06, 83.299 and 3.463 respectively, outperforming other methods and techniques as shown in their comparison.

Various popularity metrics have been used in previous studies, such as the number of likes (Purba et al. [2020]). The number of likes is more accurate on a single user or specific datasets (De et al. [2017], Zohourian et al. [2018]), whereas the engagement rate is more suitable for larger ones (C.J. Qian [2017]). Among the studies reviewed there were discretized metrics, such as popular/unpopular Zhang et al. [2018] and viral/non-viral (Deza and Parikh [2015]). Other metrics were intrinsic popularity (Ding et al. [2019]) and view count per day (Wu et al. [2017]).

Regarding commonly used datasets for this task, most studies reviewed use a custom dataset extracted from the twitter API over a specific period. Usually, tweets are gathered at the time of performing the experiment, thus reflecting the current trends and state of the community.

1.1.2.2. Topic modeling

In order to achieve a good hashtag recommendation system, one of the main tasks in our work is the development of a comprehensive topic modeling system. Topic modeling is an unsupervised machine learning technique for detecting words and phrase patterns within a set of documents, with Latent Dirichlet Allocation (LDA, Blei et al. [2003]) being one of the most common approaches. In this way the related topics for a given tweet can be retrieved, and thus finding related hashtags will be easier and more accurate.

Since the creation of the first social network the analysis of topics in Twitter and various micro-blogging messages is a research area with high and rapidly growing interest within the academic community. For example, in Krestel et al. [2009], the authors trained a specific topic model method with LDA in order to obtain a topical distribution for a collection of over 3 million tagged documents. Then, they chose the top word of the dominant topics in the tweet as their recommended tag, improving the F-score obtained with association rules in almost 20 percentage points. Also, in Zhao et al. [2021] the authors used an LDA-driven technique to identify users that were the most similar to a single user, and then recommended hashtags to the user based on the hashtags of the similar users. They use a custom dataset of over 1.5 million tweets an improve several methods such as EUCF and T-LDA (see Zhao et al. [2011]) in a hit-rate based metric.

1.1.2.3. Hashtag, mention and time recommendation system

Automatic hashtag recommendation is a novel trend that has attracted researchers' attention in recent years, as demonstrated in a survey by Amiri et al. [2021]. To the authors' knowledge, Amiri et al. is the first survey article focused on hashtag recommendation. For example, in Tariq et al. [2013] the authors proposed to establish a topic-term relationship, so that the user's perception was considered to recommend the hashtag. The dataset employed consisted of 14 million tweets collected over a period of 52 days. The best results are achieved with the DTW-based system obtaining a Mean Precision of 81. In Zangerle et al. [2013], the authors approached the challenge of recommending appropriate hashtags for hyperlinked tweets as a learning-to-rank problem and used the SVMRank model to rank and suggest suitable hashtags in each tweet. The evaluations conducted showed that this approach can provide users with suitable recommendations for hashtags reaching recall values of about 22%. As an alternative approach, Gong et al. [2015] propose to model the type of hashtag as a hidden variable into their DPMM-based method (Dirichlet Process Mixture Models). Unlike other similar research works the dataset was collected from Sina Weibo1, which provides a Twitter-like service and is one of the most popular one in China. The method proposed achieved a Precision of 50.2%, a Recall of 44.2% and a F1 of 47%.

There are also various studies that have used a deep learning approach to solve the hashtag recommendation problem. For example, Li et al. used a Long Short-Term Memory architecture (LSTM) to encode the intrinsic (syntactic or semantic) relations between sentences into micro-blog vectors, and these vectors were used as features to classify tweets into various hashtags. The incorporation of topical attention mechanism gives more than 7.4% improvement in F1 score compared with standard LSTM method. Alternatively, Gong and Zhang proposed an attention-based convolutional neural network, which incorporated a local attention channel and a global channel for hashtag recommendation.

Using information retrieval approaches to tackle different recommendation tasks has been a fertile area of research. Chen et al. study people recommendations designed to help users find known, offline contacts and discover new friends on social networking sites (Chen et al. [2009]). In (Wang et al. [2013]) the whom-to-mention problem is formulated as a ranking problem, and several new challenges are addressed using the post context, which are not well studied in the traditional information retrieval tasks. As an alternative approach, in Gui et al. [2019] a novel cooperative multi-agent approach to mention recommendation is proposed, which incorporates dozens of more historical tweets than earlier approaches. To evaluate the effectiveness of the proposed model an existing dataset was expanded, from five tweets per user to 50 tweets per user from Twitter. Experimental results show that the proposed method can significantly improve performance compared to those of other baseline approaches.

To the authors' knowledge, there is a gap in research when it comes to studying the appearance of recurrent topics depending on the time of day. The idea we pursue is to study the possible impact these kinds of variables may have in the popularity gain area. To this end, we aim to study the impact that recurrent topics and hashtags have in Twitter and use this information to extend the recommendation suite.

1.1.2.4. Paraphrasing tweets

While we could not find any studies related to paraphrasing and re-structuring tweets with the aim of increasing their popularity, we did find some tangential works associated with this task. A recent survey by R. Dale [2020] explores the current trends and techniques for automatic text generation, in which it is clear that deep models have a leading role in this data-driven era. The evolution of the field of Natural Language Processing has undoubtedly had a big impact in many state-of-the-art algorithms for automatic text generation. From all of them we emphasize GPT-3, which is a celebrated autoregressive language model that uses deep learning to produce human-like text. It was developed by OpenAI in 2020, and it has since had massive success in solving many open tasks in this area of research (Floridi et al., [2020]).

An example of these kinds of tools is Quillbot (https://quillbot.com/), which is a general-purpose framework for rephrasing and generating short paragraphs.

Bibliography

- H. Tankovska.Twitter: Most users by country, 2021.
- Dale, Robert. Natural language generation: The commercial state of the art in 2020. Natural Language Engineering, 2020. 26. 481-487. 10.1017/S135132492000025X.
- Floridi, L., Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. Minds & Machines30, 681–694 (2020). https://doi.org/10.1007/s11023-020-09548-1
- Eriko Otsuka, Scott A Wallace, and David Chiu. A hashtag recommendation system for twitter data streams. Computational social networks, 3(1):1–26,2016.
- Tao Gui, Peng Liu, Qi Zhang, Liang Zhu, Minlong Peng, Yunhua Zhou, and Xuanjing Huang. Mention recommendation in twitter with cooperative multi-agent reinforcement learning. InProceedings of the 42nd International ACMSIGIR Conference on Research and Development in Information Retrieval, pages 535–544, 2019.
- Beidou Wang, Can Wang, Jiajun Bu, Chun Chen, Wei Vivian Zhang, DengCai, and Xiaofei He. Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems. InProceedings of the 22ndinternational conference on World Wide Web, pages 1331–1340, 2013.
- Jiang Zhao and Man Lan. Ecnu: Leveraging word embeddings to boost performance for paraphrase in twitter. InProceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pages 34–39, 2015.
- Eva Lahuerta-Otero and Rebeca Cordero-Gutíerrez. Retweet or like? That is the question. page 18.
- Ethan Pancer. The popularity and virality of political social media: hashtags, mentions, and links predict likes a. page 13
- Ganna Kostygina, Hy Tran, Steven Binns, Glen Szczypka, Sherry Emery, Donna Vallone, and Elizabeth Hair. Boosting Health Campaign Reach and Engagement Through Use of Social Media Influencers and Memes. Social Media + Society, 6(2):205630512091247, April 2020.
- Nelson Oliveira. Retweet Predictive Model for Predicting the Popularity of Tweets. page 9.
- Ludwig Christian Schaupp and France Bélanger. The value of social media for small businesses. Journal of Information Systems, 28(1):187–207, 2014.
- María-José Pérez-Serrano, José-Vicente García-Santamaría, and Miriam Rodríguez-Pallares. The social media presence of ibex 35 top executives and their role as influencers. Communication & Society, 33(2):313–328, 2020.
- Allison Dormanesh, Anuja Majmundar, and Jon-Patrick Allem. Follow-up investigation on the promotional practices of electric scooter companies: content analysis of posts on instagram and twitter. JMIR public health and surveil-lance, 6(1):e16833, 2020.
- Diana Owen. New media and political campaigns. In The Oxford handbook of political communication. 2017.
- Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press,2012.

- Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. Deep learning, volume 1. MIT press Cambridge, 2016.
- Sumaya Aleidi, Dalia Alsuhaibani, Nora Alrajebah, and Heba Kurdi. ATweet-Ranking System Using Sentiment Scores and Popularity Measures.
- Shahid Iqbal, Hikmat Ullah Khan, Umar Ishfaq, Mohammed Alghobiri, and SaqibIqbal. Finding influential users in social networks based on novel features& link-based analysis.
- Radia EL Bacha and Thi Thi Zin. Ranking of Influential users based on User-Tweet bipartite graph. In2018 IEEE International Conference on Service Operations and Logistics, and Informatics
- Abdalha Ali, Muasaad Alrasheedi, Abdelkader Ouda, and Luiz Fernando Capretz. A study of the interface usability issues of mobile learning applications for smartphones from the users perspective.
- Boonchoo Jitnupong and Waraporn Jirachief pattana. Information system user interface design in software services organization: a small-clan case study. InMATEC Web of Conferences, volume 164, page 01006. EDP Sciences, 2018.
- Xiaofeng Gao, Zuowu Zheng, Quanquan Chu, Shaojie Tang, Guihai Chen, and Qianni Deng. Popularity prediction for single tweet based on heterogeneous bass model. IEEE Transactions on Knowledge and Data Engineering.
- Fatemeh Hemmatian and Mohammad Karim Sohrabi. A survey on classification techniques for opinion mining and sentiment analysis.
- Matías Gastón Silva, Martín Ariel Domínguez, and Pablo Gabriel Celayes. Analyzing the retweeting behavior of influencers to predict popular tweets, with and without considering their content.
- V. Vijeya Kaveri and V. Maheswari. A framework for recommending health-related topics based on topic modeling in conversational data (twitter).
- Joaquín Osorio-Arjona, Jiri Horak, Radek Svoboda, and Yolanda García-Ruíz. Social media semantic perceptions on Madrid metro system: Using twitter data to link complaints to space.
- Yusuf Mucahit Cetinkayȧ Ismail Hakkı Toroslu, and Hasan Davulcu. Developing a twitter bot that can join a discussion using state-of-the-art architectures.
- Saswata Roy, Brijendra Kumar Suman, Joydeep Chandra, and Sourav Ku-mar Dandapat. Forecasting the future: Leveraging RNN based feature concatenation for tweet outbreak prediction.
- Chengcheng Gou, Huawei Shen, Pan Du, Dayong Wu, Yue Liu, and Xueqi Cheng. Learning sequential features for cascade outbreak prediction. Knowledge and Information Systems,
- Zekarias T. Kefato, Nasrullah Sheikh, Leila Bahri, Amira Soliman, Alberto Montresor, and Sarunas Girdzijauskas. Cas2vec: Network-agnostic cascade prediction in online social networks.
- Jiachen Sun and Peter Gloor. Towards re-inventing psychohistory: Predicting the popularity of tomorrow's news from yesterday's twitter and news feeds.
- Rishabh Misra. News category dataset, 06 2018.
- Kristo Radion Purba, David Asirvatham, and Raja Kumar Murugesan. An analysis and prediction model of outsiders percentage as a new popularity metric on instagram.
- Shaunak De, Abhishek Maity, Vritti Goel, Sanjay Shitole, and Avik Bhattacharya. Predicting the popularity of instagram posts for a lifestyle magazine using deep learning.
- Alireza Zohourian, Hedieh Sajedi, and Arefeh Yavary. Popularity prediction ofimages and videos.

- M.A. Penza C.M. Ferri C.J. Qian, J.D. Tang. Instagram popularity prediction via neural networks and regression analysis,.
- Zhongping Zhang, Tianlang Chen, Zheng Zhou, Jiaxin Li, and Jiebo Luo. How to become famous: Post popularity prediction with dual-attention.
- Arturo Deza and Devi Parikh. Understanding image virality.
- Keyan Ding, Kede Ma, and Shiqi Wang. Intrinsic image popularity assessment.
- Bo Wu, Wen-Huang Cheng, Yongdong Zhang, Qiushi Huang, Jintao Li, and TaoMei. Sequential prediction of social media popularity with deep temporal con-text networks.
- David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation.
- Ralf Krestel, Peter Fankhauser, and Wolfgang Nejdl. Latent dirichlet allocation for tag recommendation.
- Feng Zhao, Yajun Zhu, Hai Jin, and Laurence T. Yang. A personalized hashtag recommendation approach using LDA-based topic model in microblog environment.
- Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. Comparing twitter and traditional media using topic models.
- Babak Amiri, Ramin Karimiang hadim, Navid Yazdanjue, and Liaquat Hos-sain. Research topics and trends of the hashtag recommendation do-main.
- Amara Tariq, Asim Karim, Fernando Gomez, and Hassan Foroosh. Exploiting topical perceptions over multi-lingual text for hashtag suggestion on twitter.
- Eva Zangerle, Wolfgang Gassler, and G¨unther Specht. On the impact of text sim-ilarity functions on hashtag recommendations in microblogging environments.
- Yeyun Gong, Qi Zhang, and Xuanjing Huang. Hashtag recommendation using dirichlet process mixture models incorporating types of hashtags.
- Yang Li, Ting Liu, Jing Jiang, and Liang Zhang. Hashtag recommendation with topical attention-based LSTM.
- Yeyun Gong and Qi Zhang. Hashtag recommendation using attention-based convolutional neural network.
- Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. Make new friends, but keep the old: recommending people on social networking sites.

1.1.3. Justification of the proposal

1.1.3.1. Project opportunity

The opportunity of the project is based upon the following pillars:

1. Nowadays, social networks and micro-blogging sites have undoubtedly become part of our everyday life. Among others, Twitter is one of the platforms with the highest engagement capacity, which has developed into an influential networking and knowledge gathering medium used by millions of people all over the world to share experiences in real time. However, it can sometimes be difficult to choose an appropriate text to post that accurately reflects the intended meaning of the message and at the same time makes the tweet have a meaningful impact, measured in the quantity of *retweets*, *likes* and interactions. All in all, we propose an innovative approach tries to reduce the difficulty to choose an appropriate post by recommending hashtags, mentions, publication time and paraphrasing alternatives that will make the tweet potentially be trending.
2. Machine learning is a fast-growing discipline that is highly determining the advance of crucial areas like big data analytics. The challenge is to provide machine learning models and solutions strongly principled from a theoretical point of view and able to be interpreted by

humans, but at the same time meeting the technological advancements and scalability issues foreseen for the next years. Also, Machine learning is recognized as a priority in Horizon Europe Program (more precisely, in cluster "Digital, industry & space" of pillar 2 of the programme), linked to large scale data analytics and decision making.

1.2. General and specific objectives

The objectives of the project are grouped into two main categories, methodological and applications. The methodological objectives cover a diverse range of well-established ML problems, where the main challenge is to develop several models that perform hashtag, date or content recommendations given different inputs. In the applications objectives, this project will apply these methodological developments to three different types of users. One of the possible targets of our tool could be public administrations. For this, we will use our tool to increase the reach of cultural activities in the city of Madrid. Another potential user will be private companies. We will test our recommendation system for a local newspaper and measure the impact it has on the number of subscribers.

IP1 will be responsible for the methodological objectives, while IP2 will oversee and coordinate the objectives centered on practical applications.

1.2.1. Methodological

Objective M1*: Define a family of explicit models for popularity prediction.*

In order to provide reliable results, the project will develop methods for measuring and predicting the popularity of a given tweet based on different factors. This generic objective will be divided into several tasks, each of them focused on using different types of information, with the intention of being able to merge all the predictive knowledge for the subsequent objectives.

*Task M1.1**:** Popularity prediction model based on Tweet information.*

The aim of this task is to create a consolidated model that can predict the impact and popularity of a given tweet before its publication, only by considering its content and not its context. For this endeavor to succeed, we also need to learn how to correctly preprocess the tweet and establish a baseline model to which we can compare the performance of our model.

*Task M1.2**:** Extend the popularity prediction model.*

When dealing with popularity prediction, there are other factors that affect the possible impact of a given tweet that are not directly related to its content. In this task, we propose to extend the predictive model from Task M1.1 to consider a wider range of contextual input parameters. These parameters include, but are not limited to, the number of followers of the Twitter account, the time and date of publication of the given tweet, or the history of popularity of past tweets by the same user. We think that adding this information will lead to an improvement in the performance of the predictive model, which at the same time will enhance any recommendation system that builds upon it.at the same time enhance any recommendation system that uses it.

Objective M2: *Design of a recommendation system based on popularity prediction.*

Task M2.1: Hashtag recommendation system.

We intend to develop a novel hashtag recommendation system, which will combine standard recommendation techniques with automatic text generation and popularity prediction methods to produce potentially trending hashtags.

Task M2.2: Publication date and time recommendation system.

To our knowledge, nothing has been done related to study the appearing of recurrent topics depending on the time of day. The idea we pursue is to study the possible impact these kinds of variables may have in the popularity gain subject. To this end, we aim to study the repercussion that recurrent topics and hashtags have in Twitter and use this information to extend the recommendation system form Task M2.1.

Task M2.3: Propose mentions of influential personalities related to the tweet main topic.

The '#' symbol is not the only relevant one on social networks, and more specifically on Twitter. The '@' symbol can also be used in any tweet to mention users whom the sender wants to personally alert. Since 2019 Twitter and other microblogging sites offer searching options such as "Related to a user/topic/hashtag''. We argue that if the appropriate mentions are made, it would be possible to reach a large number of users with which there is not even a followee relationship. In this regard, it has been shown that adding an automatic suggestion of a small list of candidate names can improve communication efficiency Gui et al. [2019], Wang et al. [2013]. We aim to extend our suite with recommendations of this nature.

Task M2.4: Paraphrasing the tweet for greater impact.

Apart from the mentions and hashtags, the tweet content is a critical point while posting: the way it has been written may make the impressions levels rise above unsuspected limits or, on the contrary, it could make the tweet go unnoticed. As studied in J. Zhao and M. Lan [2015], giving paraphrased versions of the tweet text can improve the popularity of the post made and enhance its impact.

The idea we pursue is to take the recommendation suite to a higher level and not only give advice over possible hashtags or mentions to enhance the tweet impact, but also provide a powerful re-paraphrasing tool that gives the proposed tweet a new syntax closer to the ones used by already popular users.

Task M2.5: Create a framework to periodically retrain the model.

Each of the recommendation systems from Tasks M2.1, M2.2, M2.3 and M2.4 relies on the existence of an updated dataset of both impactful and not impactful tweets from a wide range of topics. Given that a static and permanent dataset is incapable of holding the sufficient information to generate a recommendation system prepared for new topics, it is imperative to develop a common framework that fulfills two main objectives. Firstly, it must explore and select tweets of any current topic. Secondly, this framework must be able to retrain all the recommendation systems that rely on this dataset.

***Objective M3**: Fusion/Aggregation of recommender systems in common user interface.*
Task M3.1: Unify the user interface and the recommendation system.

As shown in Figure 1.1.2, the proposed framework will allow users to see all the recommendations provided in one screen. In this way, with a quick glance everybody will have a quick idea of the suggestions that our system provides.

Task M3.2: User interface for with predictive analysis results.

In the last task, a unified recommendation interface is proposed, but many details are hidden in it: only the hashtags, mentions, etc. are shown ordered by popularity and impact. For those curious users who want to know more about the recommendations and get details about predicted impact, an interface with graphics and more details will be provided.

Task M3.3: Create embedded web browser addon for Twitter website.

Once the system is complete, the next step consists on integrating it in Twitter. In order to achieve this, a browser addon will be created. In this way, when the user goes into Twitter and wants to post some content, but he has a doubt about which hashtag use or whom to mention, he will simply click in the STE addon and a sub window with the recommendations will appear.

Figure 1.1.2. User interface for with predictive analysis results

1.2.2. Applications

Objective A1. *Increasing the reach of cultural events organized by the Madrid City Hall*

One of the targets of this tool is public institutions. This first objective is intended to demonstrate its usefulness in the dissemination of information. For this purpose, we will disseminate the agenda of cultural activities of the Madrid City Council through its official Twitter (https://twitter.com/Madrid_Cultura) and we will evaluate the reach of it.

Task A1.1: Analyze the reach of previous years' cultural activity publications.

We will study the distribution of the impact of the publications of the cultural activities of the second half of the year. This distribution will be used to be able to divide the activities into two homogeneous groups: Group A and Group B. In this way, we will be able to promote the activities of group A in the usual way and the activities of group B will be promoted using our Tweet recommendation system.

Task A1.2: Promote the Madrid City Council's cultural events agenda and evaluate the impact of our recommendation tool

During the second half of the year, we will publish through our recommendation system the activities belonging to group B defined in the previous task. The activities corresponding to group A will be published in the usual way, through the knowledge of digital marketing experts.

In order to evaluate the impact of our system, we will compare the impact of group A activities both with the impact of previous years and with the impact of group B activities.

Objective A2. *Increasing the number of subscribers to a local newspaper*

The private businesses are another target of our tool. To that end, we will be collaborating with the newspaper AlcalaHoy (https://twitter.com/AlcalaHoy1), a local newspaper of Alcalá de Henares. We will use our tool to publish the news during half a year and study the effect of our tool from the perspective of the newspaper's subscriber base.

Task A2.1: Develop a model that predicts the baseline time evolution of newspaper subscribers.

In order to measure the impact of our tool, another task will be to develop a baseline model of temporal evolution. To do so, we will develop and test a deep learning model using LSTM networks that predicts the temporal evolution of newspaper subscribers. We will use as datasets the subscribers of the last 20 years and test our model with the last 2 years.

Task A2.2: Publish news with our recommendation system and measure the impact on the number of subscribers.

During the second half of the year, we will publish all the news through our tweet recommendation system. We will measure how the number of subscribers evolves and compare it with the base model previously developed. The objective of this task is to demonstrate a real use case and its direct impact on the profitability of a business.

1.3. Methodology and work plan

In this section, the methodology that is going to be followed in this project to achieve the objectives will be presented. Then, a description of the opportunity of the project will be stated, ending with the objectives of this research proposal.

In this research, the usual scientific method is going to be applied, which is composed of 5 basic steps: 1) Detection of those unsolved or partially solved problems in a specific domain. 2) Bibliographic revision to acquire a global knowledge from the state-of-the-art. 3) Proposal of solution, where new theoretical and methodological developments, algorithm design and implementations are proposed to improve the existing ones. 4) Validation of the proposal by means of an evaluation process using standard datasets and measures, and 5) Publication of the results in renowned conferences or journals.

The work plan is to start completing the objectives sequentially. The way we have planned the project, we would be building upon our previously developed modules to construct a more and more complete suite of algorithms and models to enhance the Twitter experience of users, institutions and businesses alike. However, A detailed schedule and description on the work packages and the contingency plans can be consulted in Section 1.5.

1.3.1. Hashtag Recommendation and popularity prediction

The Hashtag recommendation module will be complemented by the popularity prediction module. The high-level steps to follow in this case are:

- *Tweet preprocessing,* using NLP techniques.
- *Topic modeling*, extracting topics from the tweet using methods such as LDA.
- *Hashtag generation*. In this step we aim two generate hashtags in two different ways: using frequent words from similar tweets, and generating new ones with automatic text generation techniques.
- *Popularity prediction.* The generated hashtags will be passed to the module developed in Task M1.1 for a quantifiable popularity prediction.
- *Final hashtag ranking*, in which the hashtags from the previous steps will be refined and ordered according to their predicted popularity.

The expected pipeline can be seen on Figure 1.1.3:

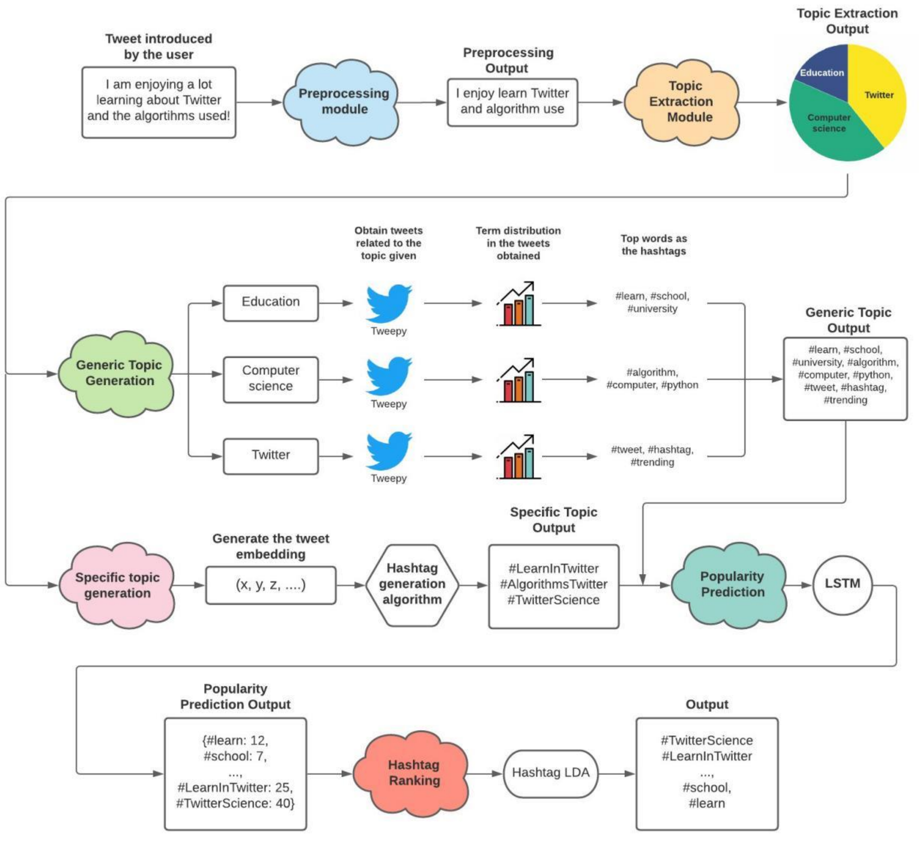

Figure 1.1.3. Hashtag Recommendation Pipeline.

1.3.2. User Recommendation

We will approach this task reusing some of the algorithms from the previous task. Our idea is to explore how well we could pair the more relevant topics of each tweets with the most influential people that talk about those topics on Twitter. We will explore graphs techniques as well as network modeling methods, to try and find the most accurate representation of the communities that revolve around a certain topic.

1.3.3. Publication Time Recommendation

In this case the methodology will be quite similar to that of the User Recommendation task. We will explore the temporal distribution of the most popular tweets, using methods such as generic

Recurrent Neural Networks and LSTMs, which have become the state of the art in time sequence modeling.

1.3.4. Paraphrasing tool

This is the final task, and the one that implies the most work. We will need to become familiar with the cutting-edge automatic text generation techniques, and then work out how to use them to model our specific problem of tweet rephrasing.

1.4. Available material resources

The group's facilities are located in a relatively modern building (built in 1999 with an extension of a new building in 2011). Laboratory space is largely available to the group for PhD students, programming staff and visitors, with room for more than 15 people, and the building has further free space for new staff if needed. The group's servers are located in a room conditioned for high performance equipment, i.e., refrigeration, fiber optics, security, etc. The group currently has several servers of different capacity acquired over the years. To keep up with growing needs for computational power and storage derived from the proposed project and the growing expectations in experiment size in the field, the project will advise the acquisition of a new computation server, the expansion of storage and the acquisition of two new desktop computers for the new PHD student and the programmer.

1.5. Schedule

***Objective** M1: Define a family of explicit models for popularity prediction.*
Responsible: Iván Cantador

Task M1.1: Popularity prediction model based on hashtags.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M1.2: Popularity prediction model based on publication date.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M1.3: Popularity prediction model based on Twitter mentions.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M1.4: Popularity prediction model based on Tweet content.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio

Deliverable 1: State of the art on popularity prediction models.
Deliverable 2: Paper describing the popularity prediction model based on hashtags.
Deliverable 3: Paper describing the popularity prediction model based on publication date.
Deliverable 4: Paper describing the popularity prediction model based on Twitter mentions.
Deliverable 5: Paper describing the popularity prediction model based on tweet content.

***Objective** M2: Design of a recommendation system based on popularity prediction.*
Responsible: Iván Cantador

Task M2.1: Hashtag recommendation system.
Responsible: A. Bellogín Kouki; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M2.2: Publication date recommendation system.
Responsible: A. Bellogín Kouki; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M2.3: Propose tweet mentions of influential personalities related to the tweet main topic.
Responsible: A. Bellogín Kouki; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M2.4: Paraphrasing the tweet for greater impact.
Responsible: A. Bellogín Kouki; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio

Deliverable 1: State of the art on Twitter recommender systems
Deliverable 2: Paper describing a hashtag Twitter recommender system.
Deliverable 2: Paper describing a publication date recommender system for Twitter.
Deliverable 4: Paper describing a user mention recommender system for Twitter.
Deliverable 5: Paper describing a *Paraphrasing* tweet tool for greater impact.

Objective *M3: Fusion/Aggregation of recommender systems in common user interface.*
Responsible: Iván Cantador

Task M3.1: Create unified recommendation system.
Responsible: L. Quijano Sánchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M3.2: User interface with predictive analysis results.
Responsible: L. Quijano Sánchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M3.3: Unify the user interface and the recommendation system.
Responsible: L. Quijano Sánchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task M3.4: Create embedded web browser addon for Twitter website.
Responsible: L. Quijano Sánchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio

Deliverable 1: Base Software open source.
Deliverable 2: The user interface with the predictive analysis result
Deliverable 3: The user interface with a unified recommendation system.
Deliverable 4: A web browser addon for Twitter website.

Objective *A1: Increasing the reach of cultural events organized by the Madrid City Hall*
Responsible: L. Quijano Sánchez

Task A1.1: Analyze the reach of previous years' cultural activity publications.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task A1.2: Promote the Madrid City Council's cultural events agenda and evaluate the impact of our recommendation tool.
Responsible: Iván Cantador; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio

Deliverable 1: Detail analysis of the impact of previous years' cultural activity publications
Deliverable 2: Evaluation of the impact of our recommendation tool in the promotion of the Madrid City Council's cultural events agenda

Objective *A2: Increasing the number of subscribers to a local newspaper*
Responsible: L. Quijano Sánchez

Task A2.1: Develop a model that predicts the baseline time evolution of newspaper subscribers.
Responsible: L. Quijano Sanchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio
Task A2.2: Publish news with our recommendation system and measure the impact on the number of subscribers.
Responsible: L. Quijano Sanchez; Team: J. Sanz Puig, P. Sánchez Pérez, E. Mena, M. Hernández Rubio

Deliverable 1: A baseline model of the newspaper subscribers time evolution
Deliverable 2: Comparison the baseline subscribers' predictions with the actual subscriber's evolution

Project Milestones

Milestone 1: The methodology tasks have advanced enough to offer methods and tools to the application tasks.
Milestone 2: All the application tasks are running.

Contingency Plan

- If Milestone 1 is not reached on time, then more members of the research team will be involved in the delayed methodological tasks.
- If Milestone 2 is not reached on time, then more members of the research team will be involved in the delayed application tasks. In addition, the possibility of an extension of the project might be considered.

Considering the number of tasks, mostly software development, in which the hired researchers are involved, in the case of the approved budget is not enough to hire them, we would reduce the degree of completion of the applications (prototypes instead of complete development) or try to finish them with the teams' own resources as those available PhD or Master students during the project period.

Chronogram

	Year 1				Year 2				Year 3				Year 4			
Tasks	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**	**T11**	**T12**	**T13**	**T14**	**T15**	**T16**
Milestones								**M1**		**M2**						
M1.1	D1			D2												
M1.2				D3												
M1.3				D4												
M1.4				D5												
M2.1					D1			D2								
M2.2								D3								
M2.3								D4								
M2.4								D5								
M3.1										D1						
M3.2										D2						
M3.3										D3						
M3.4										D4						
A1.1														D1		
A1.2																D2
A2.1														D1		
A2.2																D2

1.6. Hired Staff

The experimental work of the project involves a significant workload in the preparation and preprocessing of data, iterative software development, debugging and optimization, and the setup and parallel execution of batches of experiments, data collection processes, etc. To provide support in these tasks, we foresee hiring a PhD student with background in Computer Science -or a related degree- for the duration of the project. Apart from helping with the above tasks, the PhD student will give transversal support to all the work packages of the project with an expected maximum workload in the third year. Moreover, he or she will be immersed in a training program aimed at the development of the doctoral thesis.

2. Impacto esperado de los resultados - Expected results impact

2.1. Scientific and technological impact

The impact on the international community is ensured by the participation of prestigious international researchers in the work teams. In addition, the international visibility obtained by

journal publications will be reinforced, by attending specific international conferences of Machine Learning, Information Retrieval and Recommendation Systems, as well as general conferences on Data Science and Artificial Intelligence.

Its main impact is to advance research on the state of the art of twitter prediction and recommender systems. We plan to publish 4 papers in the area of twitter popularity prediction (Objective M1) and 4 papers in the area of recommender systems (Objective M2). Although our scope of study is focused on twitter, the developed methodology can be leveraged for different microblogging applications. In addition, our advances will be oriented in different subareas of recommender systems such as temporal recommendation, hashtags recommendation and text enhancement recommendation using natural language techniques.

2.1.1 Dissemination and internationalization

The international visibility obtained by journal publications will be reinforced, by attending specific international conferences of Machine Learning, Information Retrieval and Recommendation Systems, as well as general conferences on Data Science and Artificial Intelligence. In these events, apart from presenting the results obtained in the form of contributions, special sessions and / or tutorials will be organized (as has already been done in the past), in which the progress achieved will be announced. We will take advantage of these actions, the realization of visits / stays, the reception of international researchers, and the perfect alignment of the subject of this application with some of the priority lines of the Horizon Europe programme. All these circumstances will facilitate the participation of the researchers of these programmes to form networks and potentially create international projects.

2.2 Social and economic impact

As motivated in the introduction, social networks and microblogging sites have unquestionably been a part of our daily lives. Twitter has grown into a powerful networking and knowledge-gathering tool used by millions of people around the world to exchange real-time experiences. An important part of this project would be to prove that our recommendation system would generate a great impact in the current society.

We will prove one positive impact in the public administration in the objective A1. We will work directly with the digital marketing team of the cultural agenda of Madrid. In order to make a more efficient use of the public administrations resources it is important to have a fluid communication with their citizens. In objective A1 we will demonstrate that our tool can facilitate this communication between public administrations and citizens.

On the other hand, we will also demonstrate its commercial utility for private enterprises. Objective A2 will be responsible for demonstrating that this tool can generate a positive impact on the business of a local newspaper. We will collaborate for the first time with the local newspaper AlcalaHoy. The usefulness of this project is not limited to the journalistic sector but is applicable to different areas. Mainly to digital marketing, a booming discipline since the emergence of social networks.

Finally, the tool will be available under GNU General Public License. This way we make sure that we keep the copyright but allow the distribution of the tool in an open way. This license and the use cases exposed as objectives, will facilitate the adoption of this tool by public and private institutions and even individuals.

3. Capacidad Formativa – Training Capacity

The training capacity is guaranteed by the qualification of the team's researchers: all of them are PhD (full professors and assistant professors) and with active lines of research related to the objectives of the project. At this time this team can assume the incorporation of a new doctoral student. The researchers of the team have extensive experience in the supervision of PhD students (with doctoral thesis from 2009 on and under development).

3.1. Training plan

The training plan for the requested PhD student consists of the following points.

- The student will be encouraged to attend tutorials in conferences in the field and international summer schools (such as the European Summer School on Information Retrieval and the ACM Summer School on Recommender Systems) on a regular basis, as has been the case with many students to date. He is naturally expected to attend conferences, learn to communicate his research work –and understand the work of others– and build connections in the national and international community.
- The PhD student will be encouraged and supported in doing one or two internships with leading international research groups and/or corporations before completing his PhD thesis.
- The student will have access to the network of international and national collaborations of the team members, as well as the software and data resources of the group, and the shared computing resources, as well as the general infrastructures of the school. Finding destinations for international exchanges will be a matter of choice among a wide array of opportunities.
- The group has a long and systematic track of successful thesis in terms of quality of the dissertation, publications derived in top-tier outlets, international recognition of the graduated doctors, and professional trajectory after the PhD (see section 3.3 below). Members of the team have an active involvement in training activities for graduate students, such as summer schools and tutorials, and the elaboration of book material in the area of this project, such as: two chapters of the Recommender Systems Handbook 2nd ed. (2015), an international reference manual on recommender systems; a chapter of the book "Information Retrieval: A practical and multidisciplinary approach" (2011); the book "Collaborative recommendations: Algorithms, practical challenges and applications" (2019), including the editorial (by I. Cantador), and two chapters (by A. Bellogín and P. Castells).

3.2. PhD theses completed and ongoing in the group.

The following theses were completed in the group after January 2009. All of them were presented with the international mention.

- Rocío Cañamares Pérez. An analysis of popularity biases in recommender system evaluation and algorithms, October 3, 2019. Advisor: P. Castells
- Ignacio Fernández Tobías. Matrix factorization models for cross-domain recommendation: Addressing the cold start in collaborative filtering, January 13, 2017. Advisor: I. Cantador.
- Saúl Vargas Sandoval. Novelty and diversity evaluation and enhancement in recommender systems, April 24, 2015. Advisor: Pablo Castells
- Pedro G. Campos Soto. Recommender systems and temporal context: Characterization of a robust evaluation protocol in order to increase reliability of measured improvements, October 14, 2013. Advisors: F. Díez and I. Cantador.
- Alejandro Bellogín Kouki. Performance prediction and evaluation in recommender systems: An information retrieval perspective, November 30, 2012. Advisors: Pablo Castells and I. Cantador.
- María Ruiz Casado. Natural language processing for semiautomatic semantics extraction: Encyclopedic entry disambiguation and relationship extraction using Wikipedia and WordNet, October 16, 2009. Advisor: Pablo Castells.
- Miriam Fernández Sánchez. Semantically enhanced Information Retrieval: an ontology-based approach, April 2, 2009. Advisor: Pablo Castells.

Ongoing theses supervised by members of the project team and related to the project include:

- Javier Sanz-Cruzado Puig. Contact recommendation in social networks. Advisor: Pablo Castells. Foreseen to be finished by September 2021.
- Pablo Sánchez Pérez. Exploiting contextual information for touristic recommender systems. Advisor: Alejandro Bellogín. Foreseen to be finished by 2021
- Elisa Mena. Evaluation biases in recommender systems and information retrieval. Advisors: M. Sanderson, Y. Ren, P. Castells (joint supervision with RMIT University). Foreseen to be finished in 2023.
- María Hernández Rubio. Aspect-based user modeling and recommendation. Advisors: I. Cantador and A. Bellogín. Forseen to be finished in 2022.

3.3. Scientific-technical and formative context of the team and the institution.

Iván Cantador is Associate Professor (Profesor Contratado Doctor) at the Computer Science Department of UAM. Previously, he was Teaching Assistant and Assistant Professor at the above institution. He has made research stays at the Knowledge Media Institute of the Open University, UK (7 months), University of South-ampton, UK (3 months), University of Glasgow, UK (10 months), Free University of Bolzano, Italy (3 months), and University of Granada (1 month).

Alejandro Bellogín is Associate Professor (Profesor Contratado Doctor) at the Computer Science Department of UAM. Before his current position, he spent one year as a Postdoctoral Researcher at the Centrum Wiskunde & Informatika in Amsterdam with funding of the ERCIM program.

Lara Quijano Sánchez is Associate Professor (Profesor Ayudante Doctor) at the Computer Science Department of UAM. Before her current position, she works as a Postdoctoral Researcher at the Carlos III of Madrid University-Santander Big Data Institute - Interdisciplinarity and excellence in Big Data Analytics (IC3BS).

Appendices - Anexos

Appendix I. Budget

Description	Unitary Cost (€)	Duration	Quantity	Total cost (€)
Staff				
PhD Student	1500 €/month	4*14 months (14 payments)	1	84.000 € (*)
Application testers	100 €/month	Depending on the requirements	Depending on the requirements	Max 2500 €
Material				
Computers	1.000 €	x	4	4.000 €
Servers	1.200 €	x	5	6.000 €
Monitors	200 €	x	4	800 €
Congresses, dissemination, and training				
Workshops	Max 1000 €/Workshop	The entire life time of the project	-	Max 10.000 €
SIGIR'21: international ACM SIGIR conference on research and development in Information Retrieval	175 €	July 11-15, 2021 and expected 2022	Both IPs	Max 5.000 € (each year)
AISCA 2021: 5th International Conference on Artificial Intelligence, Soft Computing and Applications	100 €	Expected 2021	PhD Student + Both IPs	Max 5.000 € (each year)
European Conference on information retrieval	150€	Expected 2023 and 2024	PhD student + Both IPs	Max 5.000 € (each year)
Total				122.300 €

(*) This quantity is estimated based on the 2021 FPI Grants by the Spanish Ministry of Science.

Appendix II. SWOT (DAFO)

We present below an overview of the project major strengths, weaknesses opportunities and threats.

Strengths	Opportunities
• Broad knowledge of cutting-edge techniques in recommendation systems and machine learning in general. • A comprehensive survey of the state of the art in Twitter recommendation and popularity prediction. • Working knowledge of machine learning modeling and data security regulations. • Project in accordance with the Spanish R&D+I Strategy in Artificial Intelligence.	• Possibility of receiving project funding. • Attending international conferences to promote the project and the tool. • Make use of free software concepts to build a collaborative community with more professionals involved. • Transform the knowledge of the current trends into a novel and innovative proposal.
Weakness	**Threats**
• Data privacy laws and regulations are very stringent. • There are no formal protocols in place to enforce data security standards by default and by design. • Lack of expert knowledge in probabilistic models, which are widely used in this area.	• Getting involved in some kind of violation of the General Data Protection Regulation. • The intended community of users generally shows low level of support for the tool. • Long-term project that may result in the appearance of other similar and more powerful innovations over the course of the project.