Overview of MEX-A3T at IberLEF 2020: Fake News and Aggressiveness Analysis in Mexican Spanish

Mario Ezra Aragón[a], Horacio Jarquín-Vásquez[a], Manuel Montes-y-Gómez[a], Hugo Jair Escalante[a], Luis Villaseñor-Pineda[a,b], Helena Gómez-Adorno[c], Juan-Pablo Posadas-Durán[e] and Gemma Bel-Enguix[d]

[a]*Laboratorio de Tecnologías del Lenguaje (INAOE), Mexico*
[b]*Centre de Recherche en Linguistique Française GRAMMATICA (EA 4521), Université d'Artois, France*
[c]*Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (UNAM), Mexico*
[d]*Instituto de Ingeniería (UNAM), Mexico*
[e]*Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Zacatenco (IPN), Mexico*

Abstract

This paper presents the overview of MEX-A3T 2020, the third edition of this lab under the IberLEF conference. The main purpose of MEX-A3T is to explore different methodologies and strategies related to the analysis of social media content in Mexican Spanish. This year edition focuses in the identification of fake news and the detection of aggressive tweets. For this purpose, we provided different news from verified web sources and a corpus of tweets from Mexican users.

Keywords

Fake news detection, aggressiveness detection, MEX-A3T, IberLEF

1. Introduction

The goal of the third edition of MEX-A3T is to further improve the research in NLP tasks as well as to continue pushing the computational treatment of the Mexican Spanish. As a novelty, this year's proposal introduces a new track on fake news detection and an improved corpus for the aggressive language detection track. The MEX-A3T@IberLEF2020 has the following two tracks:

Aggressiveness Detection Track: Social networks represent a significant threat to users who are exposed to many risks and potential attacks. One of such threats is aggressive comments, which can produce long-term harm to victims, in the more accurate cases they can lead to suicide. This track follows up on last year's evaluation task; it focuses on the detection of aggressive

Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020)
EMAIL: mearagon@inaoep.mx (M.E. Aragón); horacio.jarquin@inaoep.mx (H. Jarquín-Vásquez); mmontesg@inaoep.mx (M. Montes-y-Gómez); hugojair@inaoep.mx (H.J. Escalante); villasen@inaoep.mx (L. Villaseñor-Pineda); helena.gomez@iimas.unam.mx (H. Gómez-Adorno); jposadasd@ipn.mx (J. Posadas-Durán); gbele@iingen.unam.mx (G. Bel-Enguix)
ORCID: 0000-0002-8213-957X (M.E. Aragón); 0000-0000-0000-0000 (H. Jarquín-Vásquez); 0000-0002-7601-501X (M. Montes-y-Gómez); 0000-0003-4603-3513 (H.J. Escalante); 0000-0003-1294-9128 (L. Villaseñor-Pineda); 0000-0002-6966-9912 (H. Gómez-Adorno); 0000-0001-9496-1328 (J. Posadas-Durán); 0000-0002-1411-5736 (G. Bel-Enguix)

tweets in Mexican Spanish. However, for this year, the criteria for identifying aggression have been revised and a new enhanced data set has been created.

Fake News Detection Track: Fake news provide information that aims to manipulate people for different purposes: terrorism, political elections, advertisement, among others. In social networks, misinformation extends in seconds among thousands of people, so it is necessary to develop tools that help control the amount of false information on the web. Particularly, fake news detection systems aim to help users to detect and filter out potentially deceptive news. The Fake News Detection Track consists in classifying a given set of news written in Mexican Spanish between true and fake.

The remainder of this paper is as follows: Section 2 covers a brief description of the previous edition of MEX-A3T. Section 3 presents the evaluation framework used at MEX-A3T 2020. Section 4 shows an overview of the participating approaches. Section 5 reports and analyzes the results obtained by the participating teams. Finally, Section 6 presents our conclusions from this evaluation exercise.

2. MEX-A3T 2019

MEX-A3T is a forum for the analysis of social media content in Mexican Spanish. Last year, we organized the second edition of the MEX-A3T shared task [1], focusing on the problems of author profiling and aggressiveness detection. A variety of methods were proposed by the participants, comprising content-based (bag of words, word n-grams, term vectors, dictionary words, and so on), stylistic-based features (frequencies, punctuation, POS, Twitter-specific elements, slang words, and so forth), and approaches based on neural networks (CNN, LSTM, and others).

For author profiling, as a novelty of previous year's edition, it was considered the use of text and images as information sources. Our purpose was to study the relevance and complementarity of multimodal data for profiling social media users. Sadly, participants could not find an effective way to take advantage of both types of information, and did not outperform the baselines proposed.

In the case of the aggressiveness identification, the top-ranked team was UACh [2]. This team used two main kinds of features, character n-grams and word embeddings, and employed two different classifiers, an SVM and a multilayer perceptron. The main idea of their participation was the inclusion of features for giving context to the text messages, and to explore if people verbally attack differently depending on their traits and overall environment.

3. MEX-A3T 2020 Evaluation Framework

This section outlines the construction of the two used corpora, highlighting particular properties, challenges, and novelties. It also presents the evaluation measures used for both tasks.

Aggressive	Non-Aggressive
Y por que disculpa? El otro joto comenzó... Le hubieras dado dos, no mas por joto y metiche	que tiene de especial la tonta texas
Yo voy cualquiera que no seas tu @USUARIO . Viejo ladrón HDP.	@USUARIO No puedo creer que nos sigamos matando por tontas ideologías
Indios estupidos no saben pa que putas es la pasarela... una vez vi cruzar unas cabras por ahi.. no entiendo como ellas si entienden	las tontas no van al cielo es mi religión

Table 1: Aggressive and Non-Aggressive Tweets.

3.1. Aggressiveness Detection

Social networks represent a significant threat to users exposed to many risks and potential attacks. One such threat is aggressive comments, which can produce long-term harm to victims, and in some cases, they can lead to suicide. This track focuses on the detection of aggressive comments on Twitter, a topic with little study in the Ibero-American community. Participants have to develop methods to determine whether a tweet is aggressive or not. The track is challenging by the fact that tweets come from Mexican users with a variety of backgrounds and social expressions.

We built a corpus of tweets for the task of aggressiveness detection from Mexican accounts. First, we selected a set of terms that served as seeds for extracting the tweets. We used the words classified as vulgar and non-colloquial in the *Diccionario de Mexicanismos de la Academia Mexicana de la Lengua*, as well as words and hashtags identified by the *Instituto Nacional de las Mujeres*. Tweets were collected considering their geolocation. We considered Mexico City as the center and extracted all tweets that were within a radius of 500 km. We annotated the corpus using the scheme proposed in [3]. The annotation provides a specific criteria to separate a tweet from aggressive, offensive and vulgar, based on the linguistic characteristics and intent of the message. Table 1 shows some examples labeled as aggressive and non-aggressive. As can be intuited, the task of labeling aggressiveness is challenging, especially because in most cases it is necessary to interpret the message in a given context.

The collected corpus consists of more than 10 thousand tweets. For the evaluation exercise, we divided the corpus into two parts, one for training and the other for the test. Table 2 shows the distribution of this corpus. The non-aggressive class is the majority class in both partitions. For readers interested in more details, [4] describes the methodology followed for the construction of the Mexican Aggressiveness Corpus.

3.2. Fake News Detection

The Spanish Fake News Corpus is a collection of news compiled from several web sources: established newspaper websites, media companies websites, special websites dedicated to

Class	Training Corpus	Test Corpus
Not Aggressive	5222	2238
Aggressive	2110	905
Σ	7332	3143

Table 2: Mexican aggressiveness corpus: distribution of the classes.

validating fake news, websites designated by different journalists as sites that regularly publish fake news. The news were collected from January to July of 2018 and all of them were written in Mexican Spanish [5]. The assembled corpus has 971 news.

The corpus was manually labeled using two classes (true or fake), and considering the following criteria:

- A news report is true if there is evidence that it has been published in reliable sites.
- A news report is fake if there are news from reliable sites or specialized websites in the detection of deceptive content that contradict it, or if no other evidence was found about the news besides the given source.

The data collection includes true-fake news pairs of different events to have a corpus as balanced as possible. Additionally, in order to avoid topic bias, the corpus covers news from 9 different topics: Science, Sport, Economy, Education, Entertainment, Politics, Health, Security, and Society. As can be seen in Table 3, the number of fake and true news is balanced; approximately 70% are used as training corpus (676 news), and 30% as the test corpus (295 news). For readers interested in more details, [5] describes the methodology followed for the construction of the Spanish Fake News corpus.

Category	**Training corpus**		**Testing corpus**	
	True	**Fake**	**True**	**Fake**
Science	32	30	14	13
Sport	45	41	21	17
Economy	18	12	6	7
Education	6	9	4	3
Entertainment	48	55	22	23
Politics	121	105	54	43
Health	16	16	7	7
Security	11	18	6	7
Society	41	52	19	22
Σ	**338**	**338**	**153**	**142**

Table 3: Spanish Fake News Corpus: distribution of the classes.

Approach	Idiap-UAM	CIMAT	Intensos	ITCG-SD	Ares	UPB	UACh	DeepMath	UMUTeam	UGalileo
Transformers	X	X				X	X			
Traditional Deep Neural Networks								X	X	X
BoW, n-grams, Stylometrics	X		X	X	X				X	

Table 4: General approach of each participating team.

3.3. Performance Measure

For both tracks, the final score corresponds to the F_1-measure for the target class, that is, fake news and aggressive messages respectively.

4. Overview of the Submitted Approaches

At this edition, eleven teams submitted one or more solutions; six teams participated in the fake news detection task, and nine participated in the aggressiveness identification task. This section presents a summary of their approaches regarding preprocessing steps, features, and classification algorithms. In Table 4 we indicate the general approach used for each team. It can be appreciated that participants used three general approaches: transformers, deep neural networks, and traditional representations like BoW and n-grams feeding a SVM classifier. Following, we briefly describe each of the participating methods.

- *Idiap and UAM Participation at MEX-A3T Evaluation Campaign* [6]
  - **Tasks:** Fake News Detection; Aggressiveness Detection.
  - **Team name:** Idiap-UAM
  - **Summary:** The authors used a Supervised Autoencoder (SAE), that is, a neural network that learns a representation (encoding) of input data and then learns to reconstruct the original input. They used three different types of features as inputs representation: word n-grams, char n-grams, and BETO encodings. The best performance was obtained when the autoencoder was fed with the combination of the three input representations.

- *Transformers and Data Augmentation for Aggressiveness Detection in Mexican Spanish* [7]
  - **Tasks:** Fake News Detection; Aggressiveness Detection.
  - **Team name:** CIMAT
  - **Summary:** The authors proposed two different strategies for the aggressiveness detection task. The first strategy consisted of an ensemble of different BETO models

(BERT models trained in Spanish) with majority and weighted voting schemes. The second strategy considered data augmentation, a technique to generate new instances from the original training data. They reported as best strategy the use of 20 ensemble models and adversarial data augmentation, where the model creates a new input for each misclassified sentence.

- *ITCG's participation at MEX-A3T 2020: Aggressive Identification and Fake News detection based on textual features for Mexican Spanish* [8]
  - **Tasks:** Fake News Detection; Aggressiveness Detection.
  - **Team name:** Intensos
  - **Summary:** The authors presented a traditional text classification approach, using a combination of binary and tf-idf text representations. They reported that their best result was using a SVM with this representation, without removing stop words.

- *TecNM at MEX-A3T 2020: Fake News and Aggressiveness Analysis in Spanish Mexican* [9]
  - **Tasks:** Fake News Detection; Aggressiveness Detection.
  - **Team name:** ITCG-SD
  - **Summary:** The authors presented a traditional machine learning approach, using a bag-of-words representation with TF and TF-IDF weights. Their best results were obtained when applying a neural network and a SVM classifier.

- *UPB at MEX-A3T 2020: Detecting Aggressiveness in Mexican Spanish Social Media Content by Fine-tuning Transformer-Based Models* [10]
  - **Tasks:** Aggressiveness Detection
  - **Team name:** UPB
  - **Summary:** The authors presented different approaches to fine-tune pre-trained Spanish, English, and multilingual transformer-based models. The best result they reported was using BETO, a BERT model trained in Spanish, but fine-tuned with the MEX-A3T aggressiveness train set and the HatEval Spanish dataset.

- *UACh at MEX-A3T 2020: Detecting Aggressive Tweets by Incorporating Author and Message Context* [11]
  - **Tasks:** Aggressiveness Detection
  - **Team name:** UACh
  - **Summary:** The authors explored the idea of using context information, such as message and author metadata. Their proposed approach has two stages. In the first stage, messages were classified considering only their content; this classification was done using BETO. Then, in the second stage, the predictions of the first stage were concatenated with the author and message metadata to form a new representation vector, which was employed by a XGBoost classifier.

- *GRU with Author Profiling Information to Detect Aggressiveness* [12]
  - **Tasks:** Aggressiveness Detection
  - **Team name:** DeepMath
  - **Summary:** The authors presented a bi-directional GRU model using words as inputs. The output of this model was combined with the predictions of gender and occupation of users obtained by a reference model, using a simple concatenation and considering a one-hot-encoding. At the end, the model considered only the gender and Sciences-Student occupation categories; the rest of the categories were discarded by a chi-squared feature selection criterion.

- *UMUTeam at MEX-A3T'2020: Towards Aggressiveness Identification in Mexican-Spanish tweets with linguistic features and word-embeddings* [13]
  - **Tasks:** Aggressiveness Detection
  - **Team name:** UMUTeam
  - **Summary:** The authors evaluated the characterization of aggressive messages through a set of linguistic attributes and sentence-embeddings. They used two types of classifiers, a support vector machine and two types of deep neural networks. Their best result was obtained by a Bi-LSTM network trained with FastText embeddings and combined with linguistic features.

- *Detecting Aggressiveness in Mexican Spanish Tweets with LSTM + GRU and LSTM + CNN Architectures* [14]
  - **Tasks:** Aggressiveness Detection
  - **Team name:** UGalileo
  - **Summary:** The authors proposed the use of two different architectures based on deep learning models. The first architecture consisted of a Bi-GRU and a Bi-LSTM networks, where the outputs are concatenated and then a prediction layer is added. For the second architecture, the authors used a Bi-LSTM and CNN network, then a concatenation and a prediction layer. Both architectures achieved similar results over the test dataset partition.

- *Ares Team: No system description paper*
  - **Tasks:** Fake News Detection
  - **Team name:** Ares
  - **Summary:** The authors proposed the use of a TF-IDF representation, combined with the capital letter ratio in the article, total number of words in the body of the article, and percentage of coincidence between the words of the body and the headline of the article. The variable selection algorithm is an F-test, and a linear algorithm with training through SGD classification

5. Experimental evaluation and analysis of results

This section summarizes the results obtained by the participants of MEX-A3T 2020, comparing and analyzing in detail the performance of their submitted solutions. For the final phase of the challenge, participants sent their predictions for the test partition, the performance on this data was used to rank them. We used the F1 over the interest class as the main evaluation measure.

For computing the evaluation scores we relied on the EvALL platform [15]. EvALL is an online evaluation service targeting information retrieval and natural language processing tasks. It is a complete evaluation framework that receives as input the ground truth and the predictive outputs of systems and returns a complete performance evaluation. In the following subsections, we report the results obtained by participants as evaluated by EvALL and an analysis of their results.

As baseline methods, we implemented two popular approaches that have shown to be hard to beat in both tasks: *i*) a classification model trained on the bag of words (BoW) representation, and *ii*) a Bi-GRU neural network. Also, we compared the systems' results against the result from *INGEOTEC*, the best performing system at the first MEX-A3T edition [16].

For both classification tasks the BoW approach was applied, in which we used all vocabulary from the corpora, removing stopwords and special characters. The size of the representation of each text was 14,913 for fake news detection, and 5,212 for aggressiveness identification; for classification we used a SVM classifier with linear kernel and $C = 1$. On the other hand, we also applied a Bi-GRU neural network in the task of aggressiveness identification. In this approach texts were pre-processed by removing stopwords, special characters, and converting all emojis to words (e.g. ☺ - 'cara sonriente'). As input features pre-trained Spanish FastText[17] embeddings were used, and a fully-connected softmax layer handle the class probabilities.

5.1. Aggressiveness detection results

Table 5 presents the results obtained by the teams in the aggressiveness detection task. For this task, we sort the teams by their F_1 results over the aggressive class. For extra analysis, we also report the accuracy, the macro F_1 and the F_1 in the non-aggressive class. The approach submitted by the CIMAT team obtained the best performance, outperforming all teams, and the proposed baselines.

To analyze in more detail the participants' results, we focused on the analysis of the complementariness and diversity of their predictions. To measure the complementarity, we used the Maximum Possible Accuracy (MPA) metric, which is defined as the quotient of the correctly classified instances over the total number of test instances. We considered an instance as correctly classified if *at least one* of the participating teams classified it correctly. On the other hand, to measure the diversity we used the Coincident Failure Diversity (CFD) metric [18], which focuses on calculating the error diversity among the participants predictions. The minimum value of this measure is 0 when all teams simultaneously predict a pattern correctly or wrongly, while the maximum value is 1, when the misclassifications are all unique.

Table 6 shows the results of applying the Maximum Possible Accuracy, and the Coincident Failure Diversity metrics over all participating teams and the different approaches in the aggressiveness identification task. From these results, it is possible to observe that the MPA

Team	Aggressive	Non aggressive	F_{macro}	Accuracy
CIMAT-1	**0.7998**	0.9195	**0.8596**	0.8851
CIMAT-2	0.7971	**0.9205**	0.8588	**0.8858**
UPB-2	0.7969	0.9107	0.8538	0.8759
UACh-2	0.7720	0.9042	0.8381	0.8651
Baseline(*INGEOTEC*)	0.7468	0.8933	0.8200	0.8498
Idiap-UAM-1	0.7255	0.8886	0.8071	0.8416
Baseline (*Bi-GRU*)	0.7124	0.8841	0.7983	0.8348
Idiap-UAM-2	0.7066	0.8953	0.8010	0.8451
UACh-1	0.7062	0.8861	0.7961	0.8358
DeepMath-1	0.7001	0.8544	0.7773	0.8040
DeepMath-2	0.6957	0.8537	0.7747	0.8024
Baseline (*BoW-SVM*)	0.6760	0.8780	0.7770	0.8228
UMUTeam-2	0.6727	0.8706	0.7716	0.8145
Intensos-1	0.6619	0.8752	0.7686	0.8177
UMUTeam-3	0.6516	0.8771	0.7644	0.8183
UGalileo-2	0.6388	0.8208	0.7298	0.7604
UGalileo-1	0.6387	0.8430	0.7408	0.7811
ITCG-SD	0.6080	0.8820	0.7450	0.8186
UMUTeam-1	0.5892	0.8430	0.7161	0.7728
UPB-1	0.3437	0.8463	0.5950	0.7509
Intensos-2	0.2515	0.7664	0.5090	0.6440

Table 5: Results for the aggressiveness identification task

from all teams and from the teams using the different types of approaches is considerably greater than the best performance Accuracy obtained by the CIMAT team, suggesting that the participants systems and approaches are complementary to each other. In terms of the different approaches, the Transformes approach obtained the greater MPA over the BoW and traditional DNN approaches, which is consistent with the results shown in Table 5, where the top-teams obtained their best performances using transformers. The results obtained with the CFD metric, show that there is a high error diversity in the predictions of the DNN approaches, which is consistent with the performance shown by this approach using the MPA metric. On the other hand, the approach that obtained the best performance results with the MPA metric, showed less error diversity in the participants predictions.

Approach	Best Accuracy	MPA	CFD	Number of systems
All teams	0.8858	0.9965	0.3204	18
Transformers	0.8858	0.9885	0.3549	9
BoW and n-grams	0.8177	0.9287	0.4357	3
Traditional DNN	0.8040	0.9707	0.4731	6

Table 6: Comparison of MPA and CFD results between the different general approaches

As a result of applying the MPA metric, it was possible to identify those common errors across all systems. In fact, there are only 10 tweets that no system could classify correctly. All of them are aggressive tweets that were classified as non-aggressive. Below we present some of these tweets, where we can identify offenses with no vulgar or profane words, the use of out of the training vocabulary words, as well as some ironic comments.

- No se vale estar “buena” por gorda y “flaca” por anoréxica, chingenle al gym y a la dieta.
- Viendo a #AnetteMichel con ese vestido en #MasteChefMx se me hace bien gorda.
- @USUARIO te quiero meter todo mi camote.
- Las fans de odisseo se ven bien bonitas en sus fotos de twitter y estan bien feas en persona.
- #InyeccionLetalParaLGBT y para la que me ilusiono, luego me echó la culpa de que le rompi el corazón, me dejó, y volvió con su ex... ardida!

5.2. Fake News Detection results

Table 7 shows a summary of the results obtained by each team in the Fake News detection track. As in the previous section, we report the F_1 in both fake and true classes, the macro F_1, and the accuracy. We used the F_1 over the fake class to rank participants. In this task, the approach submitted by the Idiap-UAM team outperformed all the other approaches and the baselines. It can be observed that all systems achieved balanced results in both fake and true classes, however, the F_1 score of the true class is in general slightly better in almost all systems. All participated teams used a machine-learning-based approach relying on style-based features, i.e., neither team used a knowledge base or Web searching to verify the authenticity of the news.

Team	Fake	Truth	F_{macro}	Accuracy
Idiap-UAM-1	**0.8444**	**0.8688**	**0.8566**	**0.8576**
Idiap-UAM-2	0.8406	0.8599	0.8502	0.8508
Ares	0.8188	0.8151	0.8169	0.8169
CIMAT-1	0.7943	0.8117	0.8030	0.8034
Baseline (*BoW-RF*)	0.7850	0.7879	0.7864	0.7864
Intensos-2	0.7703	0.7883	0.7793	0.7797
Intensos-1	0.7597	0.7376	0.7487	0.7492
Baseline (INGEOTEC)	0.7596	0.7723	0.7659	0.7661
ITCG-SD	0.7464	0.7771	0.7617	0.7627

Table 7: Results for the fake news detection task

The analysis of the complementariness and the diversity of the predictions of the different approaches using the MPA and CFD metrics are shown in the Table 8. The table uses the following hierarchy for the participants: Transformers approach considers only the CIMAT team, BoW and n-gram approach considers Ares, Intensos and ITCG-SD teams, Hybrid approaches includes Idiap-UAM teams. The CFD for the Transformers methodologies row could not be

calculated because there is only one participant. The MPA for the row of all teams has the highest value, which means that the teams' approaches complement each other.The best systems (Idiap-UAM 1.2) obtained a lower MPA value, around 9%, compared to that of all teams. On the contrary, the systems with BoW and n-grams approach obtained an MPA value similar to that of all the teams, showing greater complementariness in the proposed approaches. The teams that implemented BoW and n-gram approaches showed a greatest diversity of errors in their predictions, this is consistent with their MPA performance and the heterogeneity of the approaches. The lowest value for the CFD score corresponds to the Idiap-UAM runs, which means that their predictions are alike, this can be explained because both runs use the same core.

Approach	Best Accuracy	MPA	CFD	No. of systems
All teams	0.8576	0.9729	0.3531	7
Hybrid (Idiap-UAM 1,2)	0.8576	0.8814	0.1615	2
Transformers (CIMAT)	0.8034	0.8034	-	1
BoW, n-grams (Intensos 1,2 + ITCG + Ares)	0.8169	0.9458	0.3835	4

Table 8: Comparison of MPA and CFD results between the different general approaches for Fake News track

The Table 9 shows the results of the F_1 score for the fake class in the different topics of the corpus. It can be observed that the Economy category is the most difficult for all the evaluated approaches. On the contrary, there were three systems that correctly classified all the instances in the Education topic, and two systems that achieved perfect scores in the the Security topic. The performance of the systems does not seem related to the number of news each topic has. Politics, Entertainment, Sport and Science are the largest topics, while Education, Health and Security are the less represented groups. However, it seems that the most difficult topic to identify was economy, although it could seem that, having more examples, could help the system to learn better.

We identified the common prediction errors across all the systems and find that there were only 8 news, 7 in the fake class that none of the approaches classify correctly. Table 10 shows the classified instances, it can be observed that the 37.5% of the missclassified news belong to the Economy category, while the 25% are included in the group of politics. The groups of society, science and health, show one entry that has not been correctly classified by any team (12% of the total).

6. Conclusions

This paper described the design and results of the MEX-A3T shared task collocated with IberLef 2020. MEX-A3T stands for *Authorship and Aggressiveness Analysis in Mexican Spanish Tweets*. Two tasks were proposed, one targeting fake news detection and the other focused on aggressiveness detection.

Regarding aggressiveness detection, this has been the third edition of the task, and this

Team	Education	Society	Science	Security	Health	Economy	Sport	Politics	Entertainment
Idiap-UAM-1	**1.00**	0.88	**0.92**	**1.00**	0.77	0.60	0.79	0.84	0.84
Idiap-UAM-2	**1.00**	0.82	0.83	**1.00**	0.77	0.60	**0.81**	**0.85**	**0.86**
Ares	**1.00**	0.88	0.83	0.86	0.86	0.60	0.74	0.82	0.83
CIMAT-1	0.86	0.81	0.74	0.86	0.86	0.60	0.79	0.83	0.76
BoW-RF	0.86	0.85	0.73	0.92	0.86	0.60	0.68	0.81	0.77
Intensos-2	0.67	**0.91**	0.79	0.93	0.71	0.44	0.65	0.77	0.78
Intensos-1	0.75	0.85	0.69	0.92	**0.88**	**0.73**	0.68	0.79	0.67
INGEOTEC	0.88	0.75	0.77	0.86	0.86	0.60	0.74	0.76	0.75
ITCG-SD	0.75	0.80	0.64	0.67	0.77	0.55	0.69	0.79	0.79

Table 9: Results for the fake news detection task

Label	Topic	Sources	Title
True	Health	El país	Barba, una moda que daña tu salud
Fake	Society	Actualidad RT	“Las puertas del infierno”: Un extraño video universitario causa ’terror’ en la Red
Fake	Science	Rey Misterios	Asteroide contra la Tierra
Fake	Economy	Alerta digital	El Gobierno de Sánchez gastará *NUMBER* millones de euros en demoler
Fake	Economy	Lamula	Se debe pagar Impuesto a la Renta por el uso de satélites de comunicación
Fake	Economy	Voz del Sur	La CIA ya conoce la fecha del próxima caída económica que podría afectar a México
Fake	Politics	Criterio Universal	Exhiben pacto Duarte-Morena
Fake	Politics	Sin embargo	Forbes afirma que Angélica Rivera está ya en la lista de mexicanos millonarios en EU

Table 10: Fake Instances Missclassified by all Systems.

year the results have outperformed the past competitions. Clearly, the use of transformers has achieved the best results, and shows the appropriateness of this method for approaching this key topic in NLP.

Although this has been the first edition of the task in fake news detection, the results that have been achieved are really promising. Contrary to the task of aggressiveness detection, the best results here have been reached by hybrid approaches, using both transformers and BoW-n-grams, or just n-grams. Traditional Deep Neural Networks have not been used in this task.

Summing up, the achievements of these tasks of the IberLef evaluation forum showed how some key topics in NLP using Spanish as a source language have experienced a great development

in recent years. Both data compilation and the use of cutting-edge methods, have placed Spanish among the languages with the most accurate applications in the area of natural language processing.

Acknowledgments

Our special thanks go to all of MEX-A3T's participants. We would like to thank CONACyT for partially supporting this work under grants CB-2015-01-257383, FC-2016-2410, CB-A1-S-27780, the Thematic Networks program (Language Technologies Thematic Network), and UNAM under PAPIIT projects IA401219, TA100520. The first author thanks for doctoral scholarship CONACyT-Mexico 654803 and the second for master scholarship CONACyT-Mexico.

References

[1] M. E. Aragón, M. Á. Álvarez-Carmona, M. Montes-y Gómez, H. J. Escalante, L. Villaseñor-Pineda, D. Moctezuma, Overview of mex-3at at iberlef 2019: Authorship and aggressiveness analysis in mexican spanish tweets, in: Notebook Papers of 1st SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Bilbao, Spain, September, 2019, p. .

[2] M. Casavantes, R. López, L. C. González, Uach at mex-a3t 2019: Preliminary results on detecting aggressive tweets by adding author information via an unsupervised strategy, in: In Proceedings of the First Workshop for Iberian Languages Evaluation Forum (IberLEF 2019), CEUR WS Proceedings, 2019, p. .

[3] M.-J. Díaz-Torres, P. A. Moran-Méndez, L. Villasenor-Pineda, M. Montes-y Gomez, J. Aguilera, L. Meneses-Lerin, Automatic detection of offensive language in social media: Defining linguistic criteria to build a mexican spanish dataset, in: Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying, 2020, p. .

[4] M. Á. Álvarez-Carmona, E. Guzmán-Falcón, M. Montes-y Gómez, H. J. Escalante, L. Villaseñor-Pineda, V. Reyes-Meza, A. Rico-Sulayes, Overview of mex-a3t at ibereval 2018: Authorship and aggressiveness analysis in mexican spanish tweets, in: Notebook Papers of 3rd SEPLN Workshop on Evaluation of Human Language Technologies for Iberian Languages (IBEREVAL), Seville, Spain, September, 2018, p. .

[5] J.-P. Posadas-Durán, H. Gómez-Adorno, G. Sidorov, J. J. M. Escobar, Detection of fake news in a new corpus for the spanish language, Journal of Intelligent & Fuzzy Systems 36 (2019) 4869–4876.

[6] E. Villatoro-Tello, G. Ramírez-de-la Rosa, S. Kumar, S. Parida, M. Petr, Idiap and uam participation at mex-a3t evaluation campaign, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[7] M. Guzman-Silverio, A. Balderas-Paredes, A.-P. López-Monroy, Transformers and data augmentation for aggressiveness detection in mexican spanish, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[8] D. Zaizar-Gutierrez, D. Fajardo-Delgado, M.-A. Álvarez Carmona, Itcg's participation at mex-a3t 2020: Aggressive identification and fake news detection based on textual features

for mexican spanish, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[9] S. Arce-Cardenas, D. Fajardo-Delgado, M.-A. Álvarez Carmona, Tecnm at mex-a3t 2020: Fake news and aggressiveness analysis in spanish mexican, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[10] M.-A. Tanase, G.-E. Zaharia, D.-C. Cercel, M. Dascalu, Upb at mex-a3t 2020: Detecting aggressiveness in mexican spanish social media content by fine-tuning transformer-based models, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[11] M. Casavantes, R. López, L.-C. González, Uach at mex-a3t 2020: Detecting aggressive tweets by incorporating author and message context, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[12] M.-G. Garrido-Espinosa, A. Rosales-Pérez, A.-P. López-Monroy, Gru with author profiling information to detect aggressiveness, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[13] J.-A. García-Díaz, R. Valencia-García, Umuteam at mex-a3t'2020: Towards aggressiveness identification in mexican-spanish tweets with linguistic features and word-embeddings, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[14] V. Peñaloza, Detecting aggressiveness in mexican spanish tweets with lstm + gru and lstm + cnn architectures, in: Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain, September, 2020, p. .

[15] E. Amigó, J. Carrillo-de Albornoz, M. Almagro-Cádiz, J. Gonzalo, J. Rodríguez-Vidal, F. Verdejo, Evall: Open access evaluation for information access systems, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 2017, pp. 1301–1304.

[16] M. Graff, S. Miranda-Jiménez, E. S. Tellez, D. Moctezuma, V. Salgado, J. Ortiz-Bejar, C. N. Sánchez, Ingeotec at mex-a3t: Author profiling and aggressiveness analysis in twitter using μtc and evomsa, in: In Proceedings of the Third Workshop on Evaluation of Human Language Technologies for Iberian Languages (IberEval 2018), CEUR WS Proceedings, 2018, p. .

[17] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Learning word vectors for 157 languages, in: Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), 2018, p. .

[18] E. K. Tang, P. N. Suganthan, X. Yao, An analysis of diversity measures, Mach. Learn. 65 (2006) 247–271. URL: https://doi.org/10.1007/s10994-006-9449-2. doi:10.1007/s10994-006-9449-2.