

JUGLAR DIAZ

SENIOR DATA SCIENTIST

I specialize in **data mining and machine learning** with **ten years** of experience in research and applied projects. I have been involved in projects related to natural language processing, user data in banking and time-series analysis. I have built, trained and deployed Neural Networks models and other machine learning algorithms like: XGBoost, SVM, Logistic Regression, etc. I have developed my work mainly using the combo tensorflow-sklearn-pandas and cloud technologies like GCP and AWS.

PERSONAL INFORMATION

Address: Beauchef 851, Santiago, Chile

Personal page: <https://users.dcc.uchile.cl/~judiaz>

LinkedIn: <https://www.linkedin.com/in/juglar-diaz>

Email: juglar.diaz@gmail.com

Github: <https://github.com/juglar-diaz>

Languages: Spanish, English(Fluent)

EDUCATION

PhD Candidate in Computer Science

University of Chile, Santiago.

August 2017 - Present

Bachelor (With Honors) in Computer Science

Universidad de Oriente, Santiago, Overall GPA: 5.16/5.00

September 2006 - July 2011

SKILLS

Software Languages

Python, R, C/C++

Technologies

tensorflow, pytorch, sklearn, pandas, Airflow, Pyspark, GCP, AWS

WORK EXPERIENCE

Auxiliary professor

University of Chile

March 2019 - January 2020

- Auxiliary professor of Data Mining, NLP and Deep Learning.

Researcher in data mining/machine learning

Center for Pattern Recognition and Data Mining (CERPAMID)

September 2011 - August 2016

- Researched on text mining algorithms for social media text data.

PROJECTS

Fraud detection: As part of a team I trained and deployed a fraud detection model to detect identity theft. This project was developed with Airflow and AWS services: Sagemaker, Glue, Athena and SNS.

Neural networks models for spatio-temporal textual data: I used recurrent neural networks and embedding models to jointly model text, time(timestamp) and space(coordinates). This project was developed using tensorflow and resulted in a published paper.

Crime incidents analysis: I combined multiple data sources to build time-series and spatial prediction models for crime incidents. I used python libraries: sklearn, nltk, pandas, matplotlib and gensim. The results of this project were published in a paper.

Language identification using feature weighting: I implemented Information-Gain, Mutual-Information, Kullback-Leibler and Chi-Squared as feature weighting models for Language Identification. This project was developed in Python and resulted in a published paper.

Correference resolution module for spanish texts: I implemented a clustering algorithms based on graphs for correference resolution. This project was developed in C++ and resulted in a published paper.

Sentiment analysis prediction models for social media posts: I used sklearn classifiers, recurrent neural networks and the Transformer architecture to build sentiment analysis prediction models for social media posts. I had to deal with class imbalance and scarce data. This project was developed using tensorflow.

PUBLICATIONS

An Integrated Model for Textual Social Media Data with Spatio-Temporal Dimensions. Juglar Diaz, Barbara Poblete and Felipe Bravo. Accepted at Information Processing and Management, 2020. <https://doi.org/10.1016/j.ipm.2020.102219>.

Car Theft Reports: a Temporal Analysis from a Social Media Perspective. Juglar Diaz and Barbara Poblete. In Companion Proceedings of the 2019 World Wide Web Conference (WWW'19 Companion), May 13–17,2019, San Francisco, CA, USA. ACM, New York, NY, USA, 4 pages.

Tweets language identification using feature weighting. Juglar Díaz Zamora, Adrian Fonseca Bruzón, Reynier Ortega Bueno. Proceedings of the Tweet Language Identification Workshop (TweetLID 2014), Vol. 1228. Arkaitz Zubiaga, Iñaki San Vicente, Pablo Gamallo, José Ramom Pichel, Iñaki Alegria, Nora Aranberri, Aitzol Ezeiza, Víctor Fresno (Eds). pp. 30 - 34. Septiembre, 2014. ISSN: 1613-0073.

User influence in tweets classification. Juglar Díaz Zamora, Adrian Fonseca Bruzón, Reynier Ortega Bueno. Proceedings of the VII International Conference of Electrical Engineering. pp. 1 - 4. Jun, 2014. ISBN: 978-959-207-529-0.

A Coreference Resolution Module for Spanish Texts. Juglar Díaz-Zamora y Yunior Ramírez Cruz. Proceedings of the RECPAT 2011. pp. 1 - 10. 2011. ISBN: 978-959-250-658-9.

CERTIFICATIONS

Professional Google Cloud Data Engineer, Professional Google Cloud Architect and Associate Google Cloud Engineer.

AWARDS

Google Latin American Research Award (PhD-2019).

Conicyt National Doctoral Scholarship (PhD-2016).