

DANAE SANCHEZ VILLEGAS

PhD Student | Speech and Language Technologies CDT

@ danaesavi@gmail.com

☎ +44 7517605321

in linkedin.com/in/danae-sanchez-villegas/

🐙 github.com/danaesavi

EDUCATION

PhD Speech and Language Technologies CDT

University of Sheffield. Expected Graduation: Sep. 2023

📅 Sep. 2019 – Present

📍 Sheffield, UK

- Areas of interest: NLP for social media, sentiment analysis, topic modelling, and multimodal machine learning techniques.

MSc. Advanced Computer Science (Distinction)

University of Sheffield.

📅 Sep. 2018 – Sep. 2019

📍 Sheffield, UK

Thesis: Machine Comprehension using Commonsense Knowledge, Sep. 2019

- Two distance-based sliding window approaches (DSWE), and a neural network method (AttNN) were implemented to solve a reading comprehension task. The models were enhanced with knowledge from external sources such as word embeddings, knowledge graph relations, and sentiment polarity values. The best overall accuracy (47.62%) was obtained by the DSWE model, whereas the AttNN obtained the highest accuracy (63.6%) on the commonsense knowledge type questions.

B.S. Computer Engineering

Instituto Tecnológico Autónomo de México. GPA. 8.91/10

📅 Aug. 2012 – Dec. 2016

📍 Mexico City, Mexico

Thesis: Time Series Modeling for Stock Market Prediction Using Machine Learning Techniques, May 2017

- Stock market price prediction using an ANN, SVM and RNN. Two criteria were used to compare the models: the prediction and movement's direction of the price. The RNN had the best performance with a MSE of 2.14 +/- 0.42.

EXPERIENCE

Data Scientist

deep_dive (Deep Dive Data Science S.A.P.I. de C.V.)

📅 Mar. 2017 – Sep. 2018

📍 Mexico City, Mexico

Social Media and News Analysis | Python, Elastic Search, APIs, NLP, GeoPandas, Folium, Bokeh

- Collected posts from Twitter, Facebook, and newspapers to analyze the characteristics of 3 areas in Mexico for a real state consultancy firm. We presented an interactive map and a dashboard to show the analysis of this information. This helped the firm to understand relevant features the properties should have to attract potential investors.

Summer Research Internship

Computer Intelligence Laboratory (Nanyang Technological University)

📅 Jun. 2016 – Aug. 2016

📍 Singapore

- Worked alongside SenticNet team on sentiment analysis. Analyzed results for a sentiment model using data translated from English to Spanish and found that wrong translations negatively affected the classification (positive, negative or neutral). Analyzed patterns on these events and suggested changes to improve the final results.

SKILLS

Python, Pandas, NumPy, Scikit-learn
SQL, MongoDB, Elastic Search, Java
Machine learning, Probability, Math
Data cleaning, Data visualization
Git, command line
Design Thinking, Scrum

PROJECTS

NER | Python, NLP, Classification

- Learn a named entity recogniser (NER) using a structured perceptron. The model predicted one of 5 labels for each word in a sequence. Tested in 3 different feature sets and obtained an F1-score of 0.779 on a test dataset.

Fraud Detection Model | Python, SQL, Mongo DB, AWS, Scikit-learn, Machine Learning

- Developed a fraud detection machine learning model for a Fintech company's mobile point-of-sale system. We delivered an automatic daily report of last day's transactions with an analysis and a fraud score. This process eliminated the manual work of 2 analysts and reduced the process from 30 days to 1.

Data Collection| Python, Web scraping, Elastic Search, MongoDB

- Implemented automatic processes to collect, transform, and load data from web pages (blogs, newspapers, CRMs,) to distributed databases.

AWARDS & ACTIVITIES

- CDT Studentship, Sep. 2019 - Sep 2023
- Scholarship, Consejo Nacional de Ciencia y Tecnología (Conacyt), Sep. 2018 - Sep.2019
- Honorable Mention, Ex ITAM Research Award, Mar. 2018

LANGUAGES

- English, fluent (TOEFL iBT 105/120)
- Spanish, native language
- German, basic (ÖSD Zertificat A1)

COURSEWORK

- Machine Learning and Adaptive Intelligence
- Text Processing
- Natural Language Processing
- Speech Technology