José Oñate López

Data Scientist & Analyst

Data Scientist & Analyst with ten years of experience optimizing business processes using Python, with background applying and designing custom Machine Learning models for Anomaly Detection and NLP.

WORK EXPERIENCE

Data Scientist **May 2016 – Present**
Genesys (Remote)

- Patented unsupervised algorithm for anomaly detection applied to time series. End to end development of the Anomaly Detection component as part of a Software Product for customer environment monitoring.
- Anomaly detection in log files. Custom log file parser to automatically detect event types and parameters. The approach implemented can detect anomalies based on the time series of the number of messages of specific event per minute, anomalies on metrics extracted from logs, novelty messages and anomalous sequence of events.
- Customer Sentiment Analysis. Urgency model to classify Salesforce cases based in customer interaction feeds. Fine-tuned our own sentence-transformer model. At the end, the approach implemented was able to classify customer feeds in the following 10 topics: normal, meeting request, asking for confirmation, general request, case management, urgency, misinformation, suggestion not working, disagreement and frustration.
- Case Categorization. Approach for unsupervised topic detection in Salesforce cases using subject and problem description. A deep pre-processing model was implemented to keep only relevant sentences from customer inputs. The complete approach includes document cleaning and relevant sentence detection, document embedding using sentence-transformer, dimensionality reduction, unsupervised clustering, and clusters labeling.

Data Analyst **August 2014 – April 2016**
Gas Natural Fenosa

- BI Reporting Platform. ETL and business logic using python, and dashboards using Tableau. Data sources from Oracle database and platform DB in Postgres. Windows server deployment.
- Reading Process Optimization OMA. Web application to optimize the reading schedule process.
- Virtual Billing Analyst AVF. Web application to suggest the best solution for a billing anomaly. Classification model using Random Forest updated every month with the last 13 months of data. Linux server deployment.
- Virtual Claim Analyst AVR. Web application to suggest a response to customer claims (accepted or rejected). Classification model using Support Vector.

Project Engineer **August 2011 – July 2014**
Ingeniería y Soluciones Especializadas ISES S.A.S

- Fraud Detection. Neural Network model to detect fraudulent customers in energy service. In addition, was implemented an optimization process to rank customer by fraudulent probability and the estimated energy to be recovered.
- Slow Payer Customer Modeling. Neural Network model to choose the best recovery plan for customers with debt.

CONTACT

- Barranquilla, Colombia (Open to Remote)
- +57 3004377161
- jose.onate.lopez@gmail.com
- linkedin.com/in/jonatelo

SKILLS

Data Analysis:
- Numpy
- Pandas

Data Visualization / Dashboards:
- Plotly
- Dash – Bokeh - Streamlit
- Vega
- Jupyterlab
- Tableau

ML Modeling:
- Scikit-Learn
- Tensorflow - PyTorch
- Sentence Transformers
- NLTK – spaCy

DB Engines:
- Postgres - Oracle
- Cassandra - Redis
- ElasticSearch

Software Development:
- FastAPI - Requests
- TCP sockets - ZeroMQ
- Multiprocessing - Ray - Dask
- Arrow - Kafka - Spark
- Docker Containers
- Linux environments
- Git for version control
- Scrum methodology

EDUCATION

Universidad del Norte
Barranquilla, Colombia
- Master of Science specialized on Intelligent Systems.
  Awards: Distinguished Master's Thesis (Thesis score 5.0 / 5.0).
  September 2012
- Bachelor of Electronic Engineering.
  Awards: Yellow Oak Scholarship, earned during the entire study period.
  March 2010

OTHERS

Publications