

## Work Experience

### **Independent Consultant – Data Scientist & Big Data Consultant**

07/2017 – Present, Mexico & Austin ,Tx.

- Designed Dashboards in Python-Bokeh, both for the exploration of data and for the presentation of results.
- Implemented and developed pipelines to process data in different infrastructures and with different data sources.
- Created models for classifications and regressions problems using machine learning and deep learning for several business cases.
- Implemented NLP and NLU techniques and algorithms for customer segmentation, topics detection, Chatbot and in problems of the quality of data in customer services.

### **Santander Bank – Data Scientist & Big Data Consultant**

10/2015 – 07/2017, Mexico

- Developed a system to analyze and estimate the behavior of customer transactions in the Bank.
- Analyzed the information and estimated the probability of default and expected loss from the data of the credit office and the bank.
- Created codes in Python and Hive to process the logs of the bank's mobile application.
- Designed dashboard in Tableau-Hadoop for project management reports for mobile app.

### **CES Consultant & Battersea Consultant – Senior IT Consultant**

08/2015 – 06/2016, Mexico

- Created the codes to estimate poverty in Mexico 2016 in R, SPSS and Stata, it was a project developed for CONEVAL under CES consultancy.
- Developed the codes to monitor the communication between the applications and the server that allowed analyzing the activities among the users.

### **Softtek – IT Consultant & Statistical Consultant**

01/2012 – 06/2014, Mexico

- Built predictive models to control events with projections by day, week and month.
- Created a decision tree to facilitate the allocation of events to the service desk.
- Automated the daily event reports, which reduced the time from 1.5 hours to 10 minutes.
- Collaborated in the creation of metrics for service levels, and predicted the behavior of the VUCEM-SAT system for 2013.

---

## Education

- Bachelor of science in Physics and Mathematics, ESFM-IPN Mexico.
- Master level courses in Mathematics, Cinvestav Zacatenco-IPN Mexico.
- Research Assistant in Mathematics and Complex Systems Cinvestav and UPIITA Mexico.
- Technician in Computer Science, Center of Technology High School No.19, Mexico.

---

## Skills:

**Machine Learning & Deep Learning:** Python, Pandas, Scikit-learn, Tensorflow, Keras, Numpy, SciPy, Numba, Spacy, Lightgbm, Xgboost, Catboost, Multiprocessing, Dask.

**Big Data:** Apache Spark(ML, Sql, Streaming), Apache Hive, Apache Impala, Apache Hadoop.

**Other:** Scala, R, Matplotlib, Seaborn, Bokeh, SQL, Excel-Macros, Linux, AWK, SPSS, Tableau, Octave.

**Certification:** AWS Certified Machine Learning – Specialty / Credential ID:VSNEVC02JFBEQEG3

## Relevant Projects (more at <https://dlegor.github.io/projects.html>):

**Google Analytics Customer Revenue Prediction** (Kaggle data science competition) Ranked among the **6% top worldwide**, where I needed to analyze the data set of Google Merchandise Store customers to predict revenue per customer.

**Home Credit Default Risk**(Kaggle data science competition) I finished in the **top 12% worldwide**, of the most popular competition in Kaggle, with more than 7000 participants.

**Power Laws: Forecasting Energy Consumption** (DriveData data science competition) Ranked in the **top 4% worldwide**, the problem was to analyze and forecast many time series with different features and time windows.

**Robust PCA and Robust Deep Autoencoder over Time Series for detection of Outliers** Research on the comparison of the different techniques and algorithms for the detection of outliers.