Juan Guillermo Lobaton Galindo

jg.lobaton@hotmail.com | (617) 230 5276 | linkedin.com/in/juan-guillermo-lobaton-galindo-114867169 | Somerville, MA

SKILLS

- Programming Languages: **Python** (scikit-learn, pandas, numpy, scipy, keras, plotly, seaborn, matplotlib, tensorflow, flask, streamlit), SQL.
- **Machine Learning** : Regression (K-NN, Linear, Polynomial, Decision Tree, Random Forest, Boosting and bagging methods), Classification (K-NN, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes), Clustering (K-means, K-modes, DBSCAN, Affinity Propagation), Dimensionality Reduction (PCA, LDA), Outlier Detection (Isolation Forest, One class SVM).
- Miscellaneous: **Deep Learning** (MLP, CNN, RNN), Power BI, AWS, Spark (pyspark)
- Personal: **Proactive**, **Problem-Solving skills**, **Critical thinking**, **Time management**, Quick learner, Self-motivated, Teamwork.

PROFESSIONAL EXPERIENCE

Data Science Intern *January 2021 – April 2021*
MO Technologies Bogotá D.C., Colombia

- Implemented computer vision models to detect electronic systems in street view images and cables in images of electric towers as an approach to solve the problem of predictive maintenance. Created the baseline for the company to tackle future projects.
- Pioneered in the development of a pipeline to solve the problem of managing inventories using computer vision models on densely populated images.
- Supported the Data Engineering team on data extraction and data storage tasks on GCP services.

Undergraduate Researcher – Data Analytics *August 2020 – December 2020*
University of the Andes Bogotá D.C., Colombia

- Used Python to build an ETL pipeline to clean and process 100 GB of raw EEG data, which decreased total processing time by 10%.
- Developed unsupervised machine learning models to create a typology of epileptic spikes from EEG data. The resultant model enhanced the definition of the typology by 30%.
- Built supervised machine learning models to predict the stage of an epileptic crisis, improving the predictive power of previous models by 20%.
- GitHub link to the project: **https://github.com/jglobaton10/entrances-/blob/main/Graduation%20project.md**

EDUCATION

Masters in Data Analytics (Part Time) *August 2021 – Present*
University of the Andes *(**Ranked #1** University in Colombia)* Bogotá D.C., Colombia

- Cumulative GPA: **4.93/5.0**
- Relevant Coursework: Decision Analysis, ETL and Data Modeling, Analysis of Statistical Models.

B.E., Computer Science *August 2016 – December 2020*
University of the Andes *(**Ranked #1** University in Colombia)* Bogotá D.C., Colombia

- Relevant Coursework: Applied Data Science, Machine Learning Techniques, Business Intelligence, Probability and Statistics.

PROJECTS

Portfolio: **https://github.com/jglobaton10/Portfolio_J**

League of Legends champion classifier *November 2021*

Personal Project

Final Product (Web app deployed in AWS server): **http://ec2-18-191-142-227.us-east-2.compute.amazonaws.com/**

- Collected the dataset using web scraping techniques.
- Trained a CNN model for multi-class classification using tensorflow and keras and achieved a precision of 0.75 on the test set.
- Implemented the backend server in flask and deployed it to AWS EC2 instance.
- Built the frontend server in JavaScript and html and deployed it to AWS EC2 instance on a nginx server.

Visible infrastructure *October 2021 – November 2021*

University of the Andes Bogotá D.C., Colombia

Final Product (Power BI dashboard): **https://bit.ly/3cF8gBl**

- Gathered information from the client and turn it into analytical requirements (OLAP and Dashboards).
- Produced a multidimensional model for the database. Link to the diagram: https://drive.google.com/file/d/141KXgXu3R_blY364nTh3rRKVGOi38z35/view?usp=sharing
- Built an ETL process using pyspark for retrieving data from GitHub, transform it and load it into SQL server.
- Created a Power BI dashboard based on two of the analytical requirements identified.

Legends of Runaterra Power Analysis

Personal project *November 2021 - Present*

Ongoing product (Power BI dashboard): **https://bit.ly/30TnT5i**

- Generated the database model in SQL server.
- Implemented an ETL process using pyspark for data cleaning and processing and SQL server for data storage.
- Designed the look and feel of each of the pages of the report based on each of the regions of the game, special attention was paid to the graphic elements and color palettes to incorporate.
- Produced a Power BI report.

Steam top 100 Games *October 2021*

Personal project

Final Product (Power BI dashboard): **https://bit.ly/3cFY77s**

- Generated a Python script to clean and perform EDA on steam games dataset.
- Designed a Power BI dashboard to showcase KPIs and analysis about the performance of the top games of steam.

CERTIFICATES

- Google Data Analytics Specialization (8 courses) Coursera Jun. 2021, Certificate.
- Excel Skills for Data Analytics and Visualization Specialization (3 courses) Coursera Apr. 2021, Certificate.
- Applied Data Science with Python (5 courses) Coursera Jul. 2020, Certificate.