Edgar **Fragoso**

Data Scientist

(+52) 55-211-849-86 | edgarfragosogarcia@gmail.com.mx | edgar-omar-fragoso-garcia-69006092

Summary

I am a graduate engineer with a passion for Data Science with 9 years of experience in which I have gained valuable knowledge and a good understanding of mathematics and statistical principles through different companies. My biggest challenge has been working on my self-awareness to achieve the best communication, best make decisions and more importantly, show empathy and fulfill the perspectives of the work team.

Work Experience

SomosDNA.

SF, U.S.A

Data Scientist — *Oct. 2020 - Sep. 2023*

- **Achievement:** The implementation increased the flow of users the company increase gains in 2022.
- **Challenge:** Give to end users valuable knowledge of how genetic markers are linked with clinical condition.
- **Solution:** Given a data base with the features clinical condition and genetic markers its was implemend a *Linear Model with regularization L1 and L2* to determinate what marker are linked to determinate condition.
- **Workflow:** The prediction was automated using ECR, BATCH, EC2, S3, SQS and Cloud Watch all administration through Lambda functions.
- **Tools:** MongoDb, Docker,Sklearn,Pandas,Numpy, Plotly and custumed snippet of code made in Python.

Código46.

Cuernavaca, México

Data Scientist — *Feb. 2017 - Nov. 2020*

- **Achievement:** The automatization allowed to the company concentrate its efforts on the sales area.
- **Challenge:** Give to end user information about genetic antesesors or the probabilty to belong a particular ethnic group.
- **Solution:**. Given a database with genetic–geographic information was implemented *PCA* for determine the possible clusters and *SVM* for make the predictions, it was implemented *DBSCAN* for detect the percentage of outliers
- **Workflow:** In the Linux system was mounted Python script that makes the prediction the values are saved in SQL so the user can visualize his results through the site web .
- **Tools:** Azure, Docker, SQL, PySpark, Koalas, Sklearn,Pandas, Numpy and custumed snippet of code made in Bash and Python v3.7.

SENASICA-SAGARPA.

Tecámac, México

Data Sientist — *Sep. 2013 - Feb. 2017*

- **Achievement:**. The organization gave technical tools to farmers to resolve international controversies about the possible transmission of pathogenic through foods with the rule of the FDA the above reduces the loss of millions of dollars to Mexican farmland.
- **Challenge:** Implementing a system of surveillance that predicts the type of pathogen.
- **Solution:**. *DBSCAN* was implemented to understand the presence of different pathogenic, with *Hierarchical divisive clustering* and *Logistic regression* we can determine the relation between pathogens of the same class. On the other hand, we applied the *Minimum Spanning Tree* to make observable the entry to pathogenic to national territory from other parts of the world.
- **Workflow:** In the Linux server system was mounted python code using a CLI interface that returned a pdf with plots and the predicted value.
- **Tools:** SQL, Pandas, Numpy and custumed snippet of code made in Bash and Python v2.5.

Skills

Cloud & MLOps	Azure, AWS, Databricks
Programming	Bash,Python
ML	Linear, Classification, Clustering.
OSys & DevOPS	Linux, Docker, Git
DB	MongoDB,SQL
IDEs	Jupyter, VisualStudio.
Languages	Spanish, English

Activity

Jobies

Partial Time

Resolve Challanges — *From - Present*

- This field is growing fast there are several tools for solving troubles. I take courses to update my capabilities and apply them to other types of complex problems. For instance *XGboost or LightGBM* designed to handle large and complex data using AutoMl in Databricks for the standardization of process.