

# JURGEN ARIAS

Washington D.C. · [jurgen.as@gmail.com](mailto:jurgen.as@gmail.com) · [jurgenarias.github.io](https://jurgenarias.github.io) · [LinkedIn](#)

Teacher and mathematician turned Data Scientist who enjoys analyzing and solving complex problems. I turn complex technical concepts and insights into impactful stories for a variety of audiences.

## KEY SKILLS

Natural Problem-Solver | Efficient | Persistent | Strong Mnemonic Capabilities | Fluent in Spanish

### Technical Skills

**Data:** Python | Data Analysis | Pandas | Data Cleaning | Statistics | APIs | Git | Excel and VBA

**Web:** JavaScript | HTML | CSS | Web Scraping | BeautifulSoup | D3 | Heroku | Flask

**Visualizations:** Matplotlib | Seaborn | Tableau | Plotly | Leaflet

**Data Storage:** SQL | PostgreSQL | pgAdmin | SQLAlchemy | MongoDB | PyMongo

**Machine Learning:** Linear Regression | Logistic Regression | SVM | DBSCAN | XGBoost | Decision Trees | Random Forests | KNN | K-Means | Neural Networks | CNNs | NLP | Voting Classifier | Scikit-learn | Keras | Tensorflow

## DATA SCIENCE PROJECTS

### **Speaker Classification using audio with Neural Networks and Convolutional Neural Networks**

- ❑ Predicted speakers by using their voice. Used 13,000 voice samples from 115 different speakers. Extracted numerical features from audio files by writing functions using librosa and python
- ❑ Cleaned data with pandas and built a feed forward Neural Network model using Keras and Tensorflow to predict speakers
- ❑ Also converted audio files to images using librosa and built Convolutional Neural Networks
- ❑ Achieved a 99.8% accuracy with the feed forward Neural Network on test data
- ❑ Performed deeper analysis to find out the least amount of data to achieve a 99% accuracy finding we only needed 20% of the training data to achieve that high accuracy
- ❑ Wrote three blogs about the project on Medium getting over 50k views

### **Classification Model with NLP on Reddit Posts**

- ❑ Web scraped 150k posts from Reddit to perform a binary classification of subreddits
- ❑ Cleaned the data and used NLP with pipelines for tokenization, lemmatization, CountVectorizer, TF-IDF Vectorizer, n-grams, and analyzed stop words and patterns in the data
- ❑ Identified most commonly used words
- ❑ Built Multinomial Naive Bayes, Random Forest, Extra Trees and Logistic Regression models grid searching the best parameters
- ❑ Built a Vote Classifier to take advantage of all the models together achieving a 93% accuracy on testing data

### **Finding the Best Movies of an Actor or Director**

- ❑ Created a website that filters a CSV file from IMDB to output the best movies of a certain actor, actress or director
- ❑ Built the website using bootstrap and added styling with css while performing the coding with JavaScript using D3 and wrote functions that grab the user's input to output the desired results
- ❑ Published the live website using github pages receiving users frequently

## EXPERIENCE

**Data Boot Camp Teaching Assistant** - Trilogy Education Services with the Columbia University - New York  
May 2021 - Present

Help students during class and hold office hours. The course includes:

- ❑ Excel crash course: VBA Scripting, Statistics Modeling
- ❑ Python Data Analytics: Python, APIs, JSON, NumPy, Pandas, Matplotlib, Beautiful Soup
- ❑ Databases: SQL, NoSQL, Postgres/pgAdmin, MongoDB
- ❑ Visualization: HTML, CSS, JavaScript, Leaflet
- ❑ Advanced Topics: Tableau, Hadoop, Machine Learning

**Data Boot Camp Teaching Assistant** - Trilogy Education Services with the Johns Hopkins University - Maryland  
July 2020 - July 2021

Maintain the Gitlab repository, grade projects, hold office hours and Help students excel in a 250+ hour boot camp including:

- ❑ Data ETL: Advanced Excel with VBA, Python, Pandas, Numpy, R
- ❑ Data Scraping: JSON, API, Beautiful Soup
- ❑ Data Web: HTML, CSS, JavaScript, D3, Heroku, Flask
- ❑ Data Base: SQL, PostgreSQL, pgAdmin, SQLite, MongoDB
- ❑ Data Visualization: Matplotlib, Tableau, Plotly, Leaflet
- ❑ Data Science: Statistics, Machine Learning models, Deep Learning, Big Data, Keras, Tensorflow

**Soccer Referee** - DC Sports and Recreation Assoc. - Washington DC  
December 2013 - Present

**Mathematics Lab Supervisor** - Northern Virginia Community College - Virginia  
December 2010 - August 2013

- ❑ Hired and trained tutors, managed yearly tutoring budget
- ❑ Create reports and approve timesheets
- ❑ Briefed reports for student performance, lab usage and statistics

## EDUCATION

**General Assembly** - *Data Science Immersive*  
September 2019 - December 2019  
Applied data analysis and data visualization skills in a 500-hour immersive full time course

**Virginia Tech**  
B.S. in Mathematics  
May 2009