Sabrina Starr

Reston, Virginia, United States starr.sabrina.e@gmail.com (571)337-9656 https://www.linkedin.com/in/s-starr/

SUMMARY

A dedicated, passionate Data Scientist with 7 years total work experience and a demonstrated expertise in programming and developing predictive models using advanced machine learning techniques. Proficient in Machine Learning, Artificial Intelligence, statistical analysis, data mining, Python, SQL, Jupyter Notebooks, Flask, HTML, CSS, among other technologies.

RELEVANT WORK EXPERIENCE

Data Science Fellow
General Assembly **January 2022 – April 2022**

- Developed a professional-grade data science portfolio of 5 projects with compelling stakeholder presentations and data visualizations.
- Developed a Vector Autoregression (VAR) and Autoregressive Integrated Moving Average (ARIMA) model for multivariate time series modeling.
- Utilized cosine similarity and sparse matrix techniques to create an item-based movie recommender with the MovieLens database.
- Designed an ensemble model consisting of a Bagged Decision Tree, Random Forest, and ExtraTree model with bootstrap techniques to predict survival rate of a crash.

Business Intelligence Intern
Amazon Web Services **June 2021 - September 2021**

- Trained a full text search (FTS) machine learning algorithm to accuracy levels that led to a new initiative being approved within the AWS Federal Civilian team.
- Engineered a data-cleaning python script that transformed the output of a machine learning program from an unreadable excel file to a CSV file that was imported into a third party visualization technology and ultimately used by senior management to successfully pitch the new initiative.
- Established an internal Jupyter Notebook project using Python to wrangle the A.I. platform's excel output files into readable data.
- Developed the AWS federal department's first ever static website using HTML & CSS and hosted on S3 for potential clients to upload requests for work.
- Regularly Synthesized and presented findings from AWS DataLens (A.I. platform) to senior Amazon leadership to influence business decisions.

TECHNICAL PROJECTS

Authentic Content Analysis – Did JK Rowling take from JRR Tolkien?
- Designed two binary classification machine learning algorithms at 96% accuracy to determine the authenticity and originality of JK Rowling's novels compared to JRR Tolkien's.
- Utilized data mining techniques with Python and Reddit's Pushshift API to determine patterns and correlations in two primary data sets (one set for each author).
- Developed a Natural Language Processing (NLP) model in Python to analyze, parse, and make conclusions from the data sets.
- Established a data pipeline for the primary ML models by using Python to wrangle the data coming from Reddit's Pushshift API.

Economic Prediction - Resiliency in the U.S. Economy
- Engineered a web scraping API and utilized K Means-Clustering and DBSCAN Unsupervised Learning techniques to predict economic resiliency of the United States in post-economic catastrophes.
- Utilized Python and Jupyter Notebooks to wrangle raw data into an appropriate form for ML models to understand.
- Created ML model to predict per industry damage based on primary economic metrics such as unemployment & employment-population ratio breakdown.
- Collaborated closely with teammates to synthesize and communicate findings to a larger audience.

Autism Prediction - Using Genes to Detect the Presence of Autism
- Successfully created 94% accurate XGBoost XGBClassifier Machine Learning algorithm that predicts the presence of Autism Spectrum Disorder.
- Utilized the SFARI ASD gene database.

Which School Districts Would Benefit from an Allocation in ACT Resources
- Conducted exploratory data analysis on years of standardized testing data for the state of California.
- Answered critical business questions using a logistic regression algorithm to help leadership execute decisions based on their data.
- Demonstrated logically correct python mastery with aggregation functions, programming libraries, and advanced control flow to determine which California Counties would benefit from an increase in state allocated resources.

The Effect of the Quality and Above Ground Living Area on Real Estate
- Created and iteratively refined a linear regression machine learning algorithm to predict housing prices in Ames Iowa.
- Accurately provided business insights through reporting and presentation.

EDUCATION

B.S. Computer Science 90 + Credit Hours Completed
Western Governors University • Salt Lake City Utah • 2022

Data Science Immersive Program
General Assembly • Remote • 2022

Programming for Data Science with Python & SQL Nano-degree
Udacity • Remote • 2020

SKILLS

Data Science: Neural Networks, Natural Language Processing, Recommendation Engines, Machine Learning/ Artificial Intelligence, data analysis, predictive & prescriptive analytics, data mining, data wrangling, regression & classification modeling, supervised & unsupervised machine learning.

Developer Tools: Git, VSCode, Jupyter Notebooks, TablePlus, PgAdmin, Tableau, AWS RDS, Amazon Chimes, Amazon QuickSight, AWS DataLens.

Libraries: NumPy, pandas, Plotly, PySpark, Tensorflow, PyTorch, Keras.

Frameworks: Streamlit, Flask.

Programming: Python, SQL, HTML, CSS.