Patrick Salsbury

p.r.salsbury@gmail.com | (408) 887-4634 | San Jose, CA | LinkedIn | Website | Github

EDUCATION

University of California, San Diego | B.S. Data Science   Expected December 2024
- Probabilistic Modeling & Machine Learning, Deep Learning, Data Management, Data Analysis & Inference, Business Analytics

De Anza College | A.S. Computer Science   September 2018 - June 2021
- Data Science Fundamentals, Data Structures & Algorithms, Object Oriented Analysis and Design

Certifications: IBM Data Engineering *(Coursera),* Google Data Analytics *(Coursera)*

EXPERIENCE

Data Science & Engineering Intern | *EverCharge* - Palo Alto, CA   June 2023 - September 2023
- Identified bugs, hardware issues, and installation problems that were previously unknown affecting 5-10% of all products by performing time series analysis, inferential analysis, and anomaly detection.
- Engineered **Retool** dashboards to showcase data validation and integrity issues while optimizing **PostgreSQL** database queries.
- Published 3 different data science reports that communicated answers to stakeholders' questions and highlighted areas of concern.
- Initiated a proof-of-concept project that involved training an ML anomaly detection model to identify abnormal device behavior.

Computer Science Coach | *TheCoderSchool* - Encinitas, CA   February 2023 - March 2023
- Mentored students ranging from Elementary to High School improving their knowledge of computer science.
- Guided and supported the development of apps in **Python** and **Scratch** over a 4-6 week timeframe while simultaneously teaching the fundamentals of programming and software engineering.

Data Science Intern | *HM Electronics* - Carlsbad, CA   June 2022 - September 2022
- Analyzed customers' device telemetry data, improving production quality and customer relationships by showcasing areas of concern.
- Extracted semi-structured console log data from 100+ customer devices to ingest into a **Azure Blob Data Lake**.
- Automated the **ETL** process on 30+ million raw records of data using **PySpark** on **Databricks(Microsoft Azure)**.
- Predicted customer product device failure using **Classification** and **Clustering ML** models, isolating variable indicators for tech support.
- Led **Agile** sprint presentations to showcase findings using **Python**, **SQL**, and **Power BI** to influence decision-making of product managers and software engineers.

PROJECTS

E-commerce Recommender System
- Developed a machine learning model for an e-commerce recommender system using **Sci-kit Learn**, **H2o AutoML**, and **Tensorflow.**
- Performed exploratory data analysis, feature engineering, model training and validating, and hyperparameter tuning in order to obtain a model with an accuracy, F1-score, and ROC-AUC of 0.71, 0.72, 0.79 respectively.
- Deployed the finalized Collaborative Filtering Tensorflow model using **Flask** and **HTML** to emulate a real-world recommender system.

Real-Time Weather Data Dashboard
- Constructed dynamic **dashboards** visualizing real-time weather updates using **Tableau**, **MongoDB, Docker,** and weather APIs.
- Highlighted temperature, humidity, and other attributes using **Apache Airflow** to build an ETL **batch** processing pipeline that ingested, transformed, and loaded the data while performing various quality checks.

Popular Youtube Video Titles Generator
- Utilized Youtube APIs to store and analyze daily top 50 trending videos to generate new possible trending video titles.
- Established an automated **batch ETL pipeline** to populate an **SQLite database** using **Bash Scripting, SQL,** and **Python**.
- Generated titles using sentence linguistics, **unigram NLP models**, and **N-gram NLP models**.

Recipe Ratings Analysis and Predictor
- Analyzed 234,429 different recipe reviews given on Food.com to **discover underlying associations** and trends regarding rating.
- Concluded statistically significant results using **hypothesis testing** and after performing **exploratory data analysis.**
- Developed a **Decision Tree model** using **CV grid search** after testing models using SKLearn pipelines using **Python**.

TECHNICAL SKILLS

Languages: Python, SQL, Java
Databases: PostgreSQL, MySQL, IBM DB2, SQLite, Redshift, MongoDB
Libraries / Frameworks: Pandas, Numpy, SciKit-Learn, PySpark, SciPy, Plotly, TensorFlow, Dask, Apache Spark, Apache Airflow
Softwares: Azure, AWS, Jira (Agile), Linux, Github, Databricks, Power BI, Tableau, Docker, Retool
Other: ML Machine Learning, Supervised Learning, Optimization, Statistical Analysis, Dashboarding, Query Optimization