Rajkumar Errakutti Sivakumar

(559)-421-8915 | rajkumarerrakutti@gmail.com | LinkedIn | esr26.github.io | California, USA

WORK EXPERIENCE

***Business Analyst Intern*, Caro Nut,** California, USA **Mar 2022 - Present**

- Collaborated with the team to develop a strategic sales plan to guide market growth efforts over the coming years using data
- Cleaned & designed a relational PostgreSQL database to preserve **14 million rows** of customer data generated from the website
- Analyzed data to provide useful insights using Excel and created dashboards using Tableau, which improved sales by **12%**

***Junior Data Scientist*, Educational and Social Research Organisation**, Chennai, India **Sep 2018 - Dec 2020**

- Utilized SQL to query **4 Terabyte (TB)** of unstructured user data on AWS Cloud and cleaned using R, which reduced the storage by **11%**
- Deployed ML model using clustering machine learning algorithms to identify potential donors, which reduced **19+** man-hours per week
- Implemented supervised machine learning techniques using Python to classify the legitimate users and automated the data pipeline using Apache Airflow, which led to a **46%** increase in donors
- Conducted code reviews and collaborated with cross-functional stakeholders to revamp code review procedures

***Data Analyst Intern,* SPI Cinemas Private Limited**, Chennai, India **May 2018 - Aug 2018**

- Queried customer data using SQL and analyzed in R to derive insights on customer behavior, informing strategies that helped to increase the customer retention rate by **9%**
- Created predictive models using Python on customer data, which impacted an increase in online sales by **32%**
- Conceptualized and implemented A/B test plans to validate customer hypotheses and understand behavior patterns

TECHNICAL SKILLS

Programming Languages: Python, R, Java, SQL, C++

Tools & Tech Stack: AWS, Spark, Apache Airflow, Google BigQuery, Tableau, PyTorch, PowerBI, Hadoop, MS Excel, Git & Version Control

Familiar Work Areas & Libraries: A/B Testing, ETL, APIs, Numpy, Sci-kit learn, Pandas, Seaborn, Matplotlib, Tensorflow, Keras

EDUCATION & HONORS

California State University, Fresno **Fresno, California**

Masters in Business Administration (MBA) (**Data Analytics Specialization**) | GPA: 3.91 / 4 **Expected: Dec 2022**

- **Coursework**: Data Warehousing, Data Mining, Machine Learning algorithms, Neural Networks, Statistical Analysis, Probability, Fuzzy logic, Data Quality, Economics, Discrete Mathematics, Data Visualization, Data Manipulation
- Robert J. Piersol Scholarship (two times awarded for being outstanding student of the program)

Anna University **Chennai, India**

Bachelor of Engineering in Electronics & Communication Engineering | GPA: 3.6 / 4 **Aug 2014 - Apr 2018**

- **Coursework**: Data Structures, Image Processing & Segmentation Analysis, Time Series Analysis, Applied Mathematics, Statistics

SELECTED PROJECTS

Customer Churn Analysis using Apache Spark & pysparkML

- Developed ML pipeline for predicting churn of a customer and performed ETL on an IBM telecom dataset
- Logistic regression and random forest models were built and cross-validated to tune the model with the best parameters
- The model had with an accuracy of **81.24%** with a precision of **0.65** and a recall of **0.5057**

Personalized Medicine: Redefining Cancer Treatment

- Predicted genetic mutations based on clinical evidence and used NLP techniques like TF-IDF and Word2Vec and categorical
  variables were encoded using OneHotEncoder and Response Coding
- K-Nearest Neighbors (KNN), Logistic Regression, random forest, SVM, and Naive Bayes models were developed
- Tuned the model hyperparameters using K-Fold Cross Validation & smoothing to attain **98.9%** accuracy
- Awarded **2nd position** at Electrofocus, an annual state-level technical symposium at Anna University

Question Pair Similarity Analysis

- Implemented a real-time duplicate questions predictor on the Quora dataset using Python and identified best features
- Performed feature extraction using NLP and Fuzzy techniques and developed Logistic regression, SVM, and xgboost models
- Improved from **76.3% to 89.6%** accuracy after hyperparameter tuning and optimizing the models
- Secured **2nd runner-up** position for the best machine learning model at Anna University hackathon with **200+** participants

Apparel Recommendation System using Deep Learning

- Developed and deployed RESTful API to recommend apparel based on text semantics in the search engine
- Synthesized a model of Semantic Analysis on Neural Networks and NLP techniques like TF-IDF, Word2Vec, and AVGW2V
- Measured the recommended products with **90.4%** accuracy and compared with product images using euclidean distance