

Rajkumar Errakutti Sivakumar

(559)-290-3955 | rajerrasiva@gmail.com | California, USA

LinkedIn: linkedin.com/in/rajkumar-errakutti/ | GitHub: github.com/esr26

PROFESSIONAL EXPERIENCE

Business Analyst Intern (Part-time), Caro Nut, California, USA

Mar 2022 - Apr 2022

- Cleaned & designed a relational databases to preserve **14 million rows** of customer data generated from the website
- Analyzed data to provide useful **data-driven business insights** and created **dashboards using Tableau**, which improved sales by **12%**
- Collaborated with the **product team** to develop strategies to guide market growth efforts and helped in **decision making process**, which reduced the cost by **8%**
- Documented business processes and key metrics to track Key Performance Indicators (**KPIs**) that facilitated smooth operations

Junior Data Scientist (Full-time), Educational and Social Research Organisation, Chennai, India

Sep 2018 - Dec 2020

- Utilized SQL to query **unstructured complex large data sets** of users on AWS Cloud and cleaned using R, which reduced the storage by **11%**
- Deployed ML model using **clustering machine learning techniques** to identify potential donors, which reduced **19+** man-hours per week
- Implemented **supervised machine learning** techniques using Python to classify 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 teams** to revamp code review procedures
- **Monitored performance** of the ML model and developed strategic plans, which reduced processing time by **13%**

Data Analyst Intern (Full-time), SPI Cinemas Private Limited, Chennai, India

May 2018 - Aug 2018

- Queried customer data using SQL and analyzed in R to derive **insights on customer engagement online**, informing strategies that helped to increase the **customer retention** rate by **9%**
- Developed 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, SQL, Java, Linux

Tools & Tech Stack: AWS, Spark, Airflow, Tableau, PowerBI, Hive, Hadoop, Jupyter Notebook, MS Excel, Git & Version Control

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

Certifications: Business Analytics Specialization (Wharton), Applied Machine Learning using Python, Google Data Analytics

EDUCATION & HONORS

California State University, Fresno

Fresno, California

Masters in Business Administration (MBA) (**Data Analytics Specialization**) | GPA: 3.92 / 4

Expected: Dec 2022

- **Relevant Coursework:** Data Modeling, Data Visualization, Analytical Tools, Data Warehousing, Regression analysis, Machine Learning Algorithms, 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 | GPA: 3.6 / 4

Aug 2014 - Apr 2018

- **Relevant Coursework:** Data Structures, Statistics, Time Series Analysis, Applied Mathematics, Cloud Computing

ACADEMIC PROJECTS

Customer Churn Analysis & Prediction using Apache Spark & pyspark ML

- Developed ML pipeline for predicting customer churn and performed ETL on an IBM telecom dataset
- Logistic regression, decision trees and random forest models were built and cross-validated to tune the model with the best parameters
- Performance metrics of the model had 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, decision trees, ensemble random forest and xgboost models
- Improved from **76.3%** to **89.6%** accuracy after hyperparameter tuning and optimization of 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