

# Rajkumar Errakutti Sivakumar

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

LinkedIn: [linkedin.com/in/rajkumar-errakutti/](https://www.linkedin.com/in/rajkumar-errakutti/) | GitHub: [github.com/esr26](https://github.com/esr26)

## PROFESSIONAL EXPERIENCE

<b>Business Analyst Intern (Part-time), Caro Nut, California, USA</b>	<b>Mar 2022 - Apr 2022</b>
<ul style="list-style-type: none"><li>Cleaned &amp; designed a relational databases to preserve <b>14 million rows</b> of customer data generated from the website</li><li>Analyzed data to provide useful <b>data-driven business insights</b> and created <b>dashboards using Tableau</b>, which improved sales by <b>12%</b></li><li>Collaborated with the <b>product team</b> to develop strategies to guide market growth efforts and helped in <b>decision making process</b>, which reduced the cost by <b>8%</b></li><li>Documented business processes and key metrics to track Key Performance Indicators (<b>KPIs</b>) that facilitated smooth operations</li></ul>	
<b>Junior Data Scientist (Full-time), Educational and Social Research Organisation, Chennai, India</b>	<b>Sep 2018 - Dec 2020</b>
<ul style="list-style-type: none"><li>Utilized SQL to query <b>unstructured complex large data sets</b> of users on AWS Cloud and cleaned using R, which reduced the storage by <b>11%</b></li><li>Deployed ML model using <b>clustering machine learning techniques</b> to identify potential donors, which reduced <b>19+</b> man-hours per week</li><li>Implemented <b>supervised machine learning</b> techniques using Python to classify legitimate users and automated the <b>data pipeline</b> using Apache Airflow, which led to a <b>46%</b> increase in donors</li><li>Conducted code reviews and collaborated with <b>cross-functional teams</b> to revamp code review procedures</li><li><b>Monitored performance</b> of the ML model and developed strategic plans, which reduced processing time by <b>13%</b></li></ul>	
<b>Data Analyst Intern (Full-time), SPI Cinemas Private Limited, Chennai, India</b>	<b>May 2018 - Aug 2018</b>
<ul style="list-style-type: none"><li>Queried customer data using SQL and analyzed in R to derive <b>insights on customer engagement online</b>, informing strategies that helped to increase the <b>customer retention</b> rate by <b>9%</b></li><li>Developed predictive models using Python on customer data, which impacted an increase in online sales by <b>32%</b></li><li>Conceptualized and implemented <b>A/B test</b> plans to validate <b>customer hypotheses</b> and understand <b>behavior patterns</b></li></ul>	

## TECHNICAL SKILLS

<b>Programming Languages:</b> Python, R, SQL, Java, Linux
<b>Tools &amp; Tech Stack:</b> AWS, Spark, Airflow, Tableau, PowerBI, Hive, Hadoop, Jupyter Notebook, MS Excel, Git & Version Control
<b>Familiar Work Areas &amp; Libraries:</b> A/B Testing, ETL, APIs, Numpy, Scikit-learn, Pandas, RESTful, Matplotlib, Tensorflow, Keras
<b>Certifications:</b> Business Analytics Specialization (Wharton), Applied Machine Learning using Python, Google Data Analytics

## EDUCATION & HONORS

<b>California State University, Fresno</b>	<b>Fresno, California</b>
Masters in Business Administration (MBA) ( <b>Data Analytics Specialization</b> )   GPA: 3.92 / 4	<b>Expected: Dec 2022</b>
<ul style="list-style-type: none"><li><b>Relevant Coursework:</b> Data Modeling, Data Visualization, Analytical Tools, Data Warehousing, Regression analysis, Machine Learning Algorithms, Data Manipulation</li><li>Robert J. Piersol Scholarship (two times awarded for being outstanding student of the program)</li></ul>	
<b>Anna University</b>	<b>Chennai, India</b>
Bachelor of Engineering in Electronics & Communication   GPA: 3.6 / 4	<b>Aug 2014 - Apr 2018</b>
<ul style="list-style-type: none"><li><b>Relevant Coursework:</b> Data Structures, Statistics, Time Series Analysis, Applied Mathematics, Cloud Computing</li></ul>	

## ACADEMIC PROJECTS

<b>Customer Churn Analysis &amp; Prediction using Apache Spark &amp; pyspark ML</b>
<ul style="list-style-type: none"><li>Developed ML pipeline for predicting customer churn and performed ETL on an IBM telecom dataset</li><li>Logistic regression, decision trees and random forest models were built and cross-validated to tune the model with the best parameters</li><li>Performance metrics of the model had an accuracy of <b>81.24%</b> with a precision of <b>0.65</b> and a recall of <b>0.5057</b></li></ul>
<b>Personalized Medicine: Redefining Cancer Treatment</b>
<ul style="list-style-type: none"><li>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</li><li>K-Nearest Neighbors (KNN), Logistic Regression, random forest, SVM, and Naive Bayes models were developed</li><li>Tuned the model hyperparameters using K-Fold Cross Validation &amp; smoothing to attain <b>98.9%</b> accuracy</li><li>Awarded <b>2<sup>nd</sup> position</b> at Electrofocus, an annual state-level technical symposium at Anna University</li></ul>
<b>Question Pair Similarity Analysis</b>
<ul style="list-style-type: none"><li>Implemented a real-time duplicate questions predictor on the Quora dataset using Python and identified best features</li><li>Performed feature extraction using NLP and Fuzzy techniques and developed Logistic regression, SVM, decision trees, ensemble random forest and xgboost models</li><li>Improved from <b>76.3% to 89.6%</b> accuracy after hyperparameter tuning and optimization of the models</li><li>Secured <b>2nd runner-up</b> position for the best machine learning model at Anna University hackathon with <b>200+</b> participants</li></ul>
<b>Apparel Recommendation System using Deep Learning</b>
<ul style="list-style-type: none"><li>Developed and deployed RESTful API to recommend apparel based on text semantics in the search engine</li><li>Synthesized a model of Semantic Analysis on Neural Networks and NLP techniques like TF-IDF, Word2Vec, and AVGW2V</li><li>Measured the recommended products with <b>90.4%</b> accuracy and compared with product images using euclidean distance</li></ul>