# SUJAI KARUNAKARAN

Chicago, Illinois 60607 | 669-264-0654 | sujaikarunakaran@gmail.com
linkedin.com/in/sujai-karunakaran | github.com/sujaikarunakaran96 | sujaikarunakaran96.github.io

## SUMMARY

- Data Scientist with **3+ years** of Professional and Academic experience in the area of Data science
- Proficient in **Python, R, SQL, Power BI, Tableau, Data Mining, Machine Learning, Big Data, Reporting & Consulting**

## EDUCATION

**Master of Science in Business Analytics** – ***The University of Illinois at Chicago (GPA – 3.77/4)*** **Expected May 2021**
**Courses:** Machine Learning, Data Mining, Advanced DBMS, Big Data Analytics, Text Analytics, Healthcare Analytics, Revenue Management, Data Visualization, Machine Learning Deployment, Time Series Forecasting, Operations Management

**Bachelor of Engineering in Biotechnology** – ***PES Institute of Technology, Bangalore, India (GPA – 7.73/10)*** **May 2017**

## CORE SPECIALIZATIONS

- Data Science
- Data Analytics
- Data Visualization
- Healthcare Analytics
- Big Data
- Predictive Analytics
- Business Intelligence
- Statistical Analysis

## SKILLS

**Programming/Analytical Tools** : Python, R, SAS, Java, C++, HTML, CSS, MATLAB, Perl
**SQL/NoSQL/Big Data** : Ms SQL, Oracle, SQL Server, Hive, Hadoop, Pig, Apache Spark, Teradata
**Data Viz./BI Tools** : Tableau, Power BI, Alteryx
**Packages** : numpy, pandas, scikit-learn, mlr, MLlib, NLTK, TensorFlow, Keras, Spark ML, caret, Torch
**Techniques/Methodology** : Regression, XGBoost, HDFS, Map-Reduce, Clustering, CNN, RNN, LSTM
**Other Tools** : AWS S3, SageMaker, Redshift, Kinesis, Data Pipeline, ML Flow, Docker, Kubernetes, Git

## PROFESSIONAL EXPERIENCE

**Data Analyst – Graduate Assistant** – UIC College of Dentistry, Chicago, USA **Feb 2020 – Present**
**Tools & Technology Used:** Python, Tableau, R, Selenium, Time Series Forecasting, NLP, Topic Modelling

- Developed a time series model to successfully forecast the Yearly, Monthly, Weekly and daily usage of network drives that helped in reducing the server operating costs by 12%
- Analyzed a large unstructured dataset of IT Help Desk tickets using Natural Language Processing, Demand Forecasting and ML models to derive operational insights that helped in optimizing product assortment and reduced inventory costs by 28%
- Performed ad-hoc data analysis to understand and predict behavior trends, in order to optimize performance

**Software Engineer** – Accenture, Chennai, India **July 2017 – April 2019**
**Tools Used:** SQL, Python, HTML, HP ALM, Java, Power BI, Selenium Automation

- Owned and led SQL modules for RTD database, involving multi-million-dollar engagements in telecommunications sector and headed automation of application programming interface testing to attain utmost functional reliability of the product
- As a Quality Lead of a cross-functional team, prepared weekly and sprint reports to evaluate the standard of quality checks using HP ALM tool, interactive Power BI dashboard and effectively reduced the percentage of non-defects from 14% to 9%
- Developed a Logistic regression model to perform root cause analysis and identified the threshold of each of the components of BT's ecommerce website in Agile Test Environment with 93% accuracy that reduced the processing time by 25%
- Developed an acquisition module using SQL for Accenture that helped reduce delivery time by 50%

## ACADEMIC PROJECTS **2019 – 2020**

**Text Mining of Online Physician Reviews** – *Python, R, SQL, Selenium Web-driver, Tableau, Gensim, Topic Modelling, NLP*

- Devised an end-to-end ETL project where reviews for 850 OBGYNs in Illinois was extracted from Healthgrades and Ratemds
- Transformed the text into new dataset by using text normalization, bag-of-words approach, Anchored LDA Topic Modelling method, Sentiment Analysis and loaded the data into SQL Database for Statistical Analysis and Visualization
- Trained a Linear Regression ML model using the transformed data to determine key factors affecting Physician's star rating
- Leveraging this model, Medical Associations and the government can implement Medical policies and regulations

**Chronic Kidney Disease Screening Tool** – *Python, SPSS, Flask, AWS Fargate, Docker, Machine Learning, Feature Selection*

- Developed a screening tool to identify patients at risk for Chronic Kidney Disease with a precision of 81% which can be used by Physicians and Hospitals to reduce the costs associated with screening tests
- Performed exploratory data analysis (EDA), descriptive statistics, data wrangling before building a Logistic Regression model and Decision Tree model, created a dashboard using Flask, containerized it using Docker and deployed it in AWS Fargate

**Recommendation System** – *Python, PySpark, Spark Machine Learning, Collaborative Filtering, Matrix Factorization*

- Analyzed a large dataset of 500k user reviews from Amazon Fine Foods to recommend products to any consumer using User-based Collaborative filtering method, ALS Matrix Factorization and Popularity based method with RMSE as an evaluation metric
- The RMSE for CF and MF methods were found out to be 1.502 and 1.204 respectively. Similarly, for a user with no prior history of ratings, popularity-based recommendation system was used

**Tool to forecast stock price of S&P500 companies** – *Python, TensorFlow, Keras, CNN, LSTM, AWS S3, Dash, AWS ECS*

- Produced an interactive tool that will accurately forecast the closing price of S&P500 companies Recurrent Neural Networks (Long Short-Term Memory) with an RMSE of 17.54
- Created a Dash application, containerized the application using Docker and deployed the application on AWS integrating S3, ECS and Fargate

## ACHIVEMENTS & OTHER ACTIVITIES

- Tableau certified "**Desktop Specialist**"
- **Data Scientist** badge – Tableau
- AWS Cloud Practitioner, Udemy