Sujai Karunakaran

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

PROFILE

- Data enthusiast with **3+ years** of professional and academic experience in Software Engineering, Data Science, and Business Intelligence
- Proficient in Python, R, SQL, Power BI, Tableau, Machine Learning, Data Mining, Big Data, Project Management, Reporting & Consulting

EDUCATION

Master of Science in **Business Analytics (Data Science),** University of Illinois at Chicago **GPA – 3.77/4** **Aug 2019 – May 2021**
Courses: Machine Learning, Data Mining, Advanced DBMS, Big Data Analytics, Text Analytics, Healthcare Analytics, Time Series Forecasting, Data Visualization, Machine Learning Deployment, Revenue Management, Operations Management | Awarded **full tuition scholarship**
Bachelor of Engineering in **Biotechnology Engineering**, PES Institute of Technology **GPA – 7.73/10** **Aug 2013 – May 2017**

SKILLS

Technical/BI Tools : Python, R Programming, SAS, SPSS, Stata, Java, MATLAB, PHP, Tableau, Power BI, Salesforce, Oracle, QlikView
Database/Big Data : Microsoft SQL Server, Oracle, PL/SQL, Hadoop, Hive, Spark, PySpark, Teradata, MongoDB
Cloud : Amazon AWS (EC2, S3, ECS, Redshift, SageMaker, Lambda, Athena), IAM, Docker, Kubernetes, Snowflake
Packages : numpy, pandas, scikit-learn, Matplotlib, NLTK, PyTorch, TensorFlow, Keras, Spark ML, Gensim
Analytics : ANOVA, A/B Testing, Time Series Forecasting, Distribution Theory, Machine Learning Techniques, RNN, LSTM

CERTIFICATIONS & COURSES

Tableau Certified **Desktop Specialist**, **Neural Networks** – Coursera, AWS Associate Developer – Udemy, Tableau **Data Scientist** Bagde

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

- Analyzed **large unstructured data** of IT Help Desk using **Natural Language Processing**, **Demand Forecasting** and ML models in **Python** that optimized product assortment and reduced inventory costs by 28%
- Developed interactive **Tableau** Visualization dashboards, a time series model to successfully forecast the Yearly, Monthly, Weekly and Daily usage of network drives thereby Reducing the server operating costs by 12% with 92% accuracy using **R**
- Performed ad-hoc data analysis to understand and predicted trends

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

- Owned and led SQL modules for RTD database, involving multi-million dollar engagements in the telecommunications sector.
- Led a cross-functional team of 4, mentored them on the functional modules, **SQL**, **Tableau** and MS Excel, trained non-technical clients
- Built Tableau dashboards for data visualizations of business impact, cleaned the data obtained from HP ALM tool, created **Tableau** and **SQL** reports that reduced the percentage of non-defects from 14% - 9%
- Developed a predictive model using **Python** to predict the downtime with 93% accuracy, reducing the processing time by 25%
- ***Project Management:*** Experience in life cycle projects from requirements gathering, as-is and to-be business process mapping, design, implementation, from Sprints to UATs and go-live, support, creation of training modules. Used **Agile** frameworks

Management Consultant Intern – Prashaste, Bangalore, India **Feb 2017 – April 2017**
Tools Used: Python, R, MS Excel

- Led a client account of automobile dealership across the country, created training modules for client workforce. Involved in start-up expansion. Frequently presented market and segment analysis to the leadership team
- Identified trends by analyzing sales data using **SQL, Python** and developed KPIs to increase revenue and decrease customer churn
- Increased the click-through rate of social media links on the website using **A/B testing** by 34%

Data Analyst Intern – Grocex, Chennai, India **May 2016 – July 2016**
Tools Used: R, MS Excel

- Assisted in the study of Indian E-Commerce market research and analyzed the difference between physical & online stores using the aspect-based approach

PROJECTS

Text Mining of Physician's reviews – *Python, R, SQL, Selenium, Tableau, Gensim, Natural Language Processing, Topic Modelling*

- Extracted large unstructured text data, engineered a dataset using NLP, supervised and semi-supervised Topic modelling and loaded the data into SQL Database for Statistical Analysis and Visualization. Trained a linear regression model to determine key factors affecting physician's star rating with an F-Score of 35.625 and Std. Error of 0.668

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

- Developed a screening tool that can identify at-risk CKD patients with an accuracy of 81% with recall of 88.2% and precision of 80%. Encased the model in Flask application, containerized using Docker, and launched it live on AWS Fargate using ECS instance

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

- Analyzed a large dataset of 500K reviews from Amazon Fine Foods to recommend products to customers based on user's reviews on other products. Built an Alternative Least Square Matrix Factorization model with an RMSE value of 1.204

Stock predictor application– *Python, Keras, TensorFlow, Dash, Docker, AWS S3, AWS Fargate, RNN, LSTM*

- Developed a Dash application that can predict the stock price of companies under S&P500 Index using Long Short Term Memory technique with an RMSE value of 15.854. Containerized the application using Docker and launched the application using AWS ECS