

# Sujai Karunakaran

Chicago, IL | 669-264-0654 | [sujaikarunakaran@gmail.com](mailto:sujaikarunakaran@gmail.com)

[linkedin.com/in/sujai-karunakaran](https://linkedin.com/in/sujai-karunakaran) | [sujaikarunakaran96.github.io/](https://sujaikarunakaran96.github.io/) | [github.com/sujaikarunakaran96](https://github.com/sujaikarunakaran96)

## PROFILE

- Data Scientist with **3+ years** of professional experience looking for full-time opportunities in Data Analytics 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 – 3.8/4** **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 – 3.1 (7.73/10)** **Aug 2013 – May 2017**

## SKILLS

**Technical/BI Tools** : Python, R, SAS, SPSS, Java, MATLAB, Tableau, Power BI, Looker, Google Data Studio, Airflow, Git  
**Database/Big Data** : Microsoft SQL Server, Oracle, PL/SQL, Hadoop, Hive, Spark, Teradata, Snowflake, PySpark  
**Cloud** : Amazon AWS (S3, ECS, Redshift, SageMaker, Lambda), Azure Cloud Service, Google BigQuery  
**Packages** : numpy, pandas, scikit-learn, Matplotlib, NLTK, PyTorch, TensorFlow, Keras, Spark ML, Gensim  
**Analytics** : ANOVA, A/B Testing, Predictive models, Data Warehousing, ML models, NLU, NLP, MLOps

## CERTIFICATIONS & COURSES

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

## PROFESSIONAL EXPERIENCE

**Data Science Intern** – Cloudbakers, Chicago, USA **Jan 2021 – Present**

- Developing a web-based application that will provide detailed insights about health of any Github repository.
- Feature engineering a pipeline using **Azure Data Factory** that will query from **Google BigQuery**, trains **Multivariate LSTM** model to accurately predict the health of the repository using **Azure Databricks** and visualize using **Power BI Embedded**.

**Data Scientist** – UIC College of Dentistry, Chicago, USA **Feb 2020 – Present**

- Developed a model to perform an automated audit of support tickets using **large unstructured data** by leveraging **Demand Forecasting**, **Topic Modeling**, and **Natural Language Processing** in **Python** that optimized product assortment and reduced inventory costs by 28%.
- Designed interactive **Power BI** 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**.
- Performing ad-hoc data analysis to understand and predict trends using **ML models**.

**Data Analyst** – Accenture, Chennai, India **July 2017 – April 2019**

- Owned and led **SQL** modules for the RTD database, created and managed the database objects including tables, indexes and **SSIS workflows**.
- Led a cross-functional team of 4, mentored them on the functional modules, **SQL**, **Tableau** and MS Excel, and trained non-technical clients.
- Created timely reports by Integrating **SQL** database with **Tableau** to visualize health of the product, business impact and to track **KPIs** that increased the product efficiency by from 86% - 91%.
- Optimized the product by creating 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 (**SDLC**).

**Management Consultant Intern** – Prashaste, Bangalore, India **Feb 2017 – April 2017**

- 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**

- Assisted in the study of Indian E-Commerce market research and analyzed the difference between physical & online stores using the aspect-based approach by developing statistical models using **R**.

## PROJECTS

**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.

**Visualizing delay statistics of Envoy Air** – *Tableau, Python, Snowflake, Excel, Statistical Analysis, LOD expressions*

- Designed interactive dashboards and stories to visualize On-time/Delay Statistics and profitable routes. Stored data in Snowflake and integrated it with Tableau to create visualizations.

**Propensity to Respond model** – *Python, Tableau, Machine Learning, Feature Selection, User Behavioral Analysis, SMOTE*

- Developed a propensity to respond model that will predict the likely user behavior and customer churn rate with a Micro-Average F1 score of 1.00 using ML models such as Random Forest, XGBoost, Ridge Regression and model stacking.