

VAISHNAVII PONNUSAMY PARAMASHIVAM

Chicago, IL | (312)934-8621 | vaishnavii.pp@gmail.com
[linkedin.com/in/vaishnavii-pp/](https://www.linkedin.com/in/vaishnavii-pp/) | [vaishnavii28.github.io/](https://github.com/vaishnavii28) | github.com/vaishnavii28

SUMMARY

- Skilled resource with 4+ years of industry experience in the areas of Data Science, Data Engineering and Data Analytics seeking full time opportunities as a Data Scientist/ Data Analyst and other related roles.
- Proficient in Python, R, SQL, Big Data, Power BI, Tableau, Advanced Statistics, Machine Learning, Cloud Computing, Reporting and Consulting.

EDUCATION

Master of Science in Business Analytics (Data Science), University of Illinois at Chicago GPA – 3.61/4 **Aug 2019 – May 2021**
Courses: Advanced Statistics, Data Mining, Machine Learning, Advanced DBMS, Big Data, Time Series & Forecasting, Healthcare Analytics, ML Deployment, Operations Management, Supply Chain Management, Analytics Strategy & Practice | Awarded **merit scholarship** for 3 semesters
Bachelor of Engineering in Computer Science Engineering, Anna University GPA – 8.09/10 **Aug 2012 – May 2016**

SKILLS

Technical/BI Tools | Python, RStudio, Java, C, PHP, JavaScript, VBA, Tableau, MS Power BI
SQL/NoSQL/Big Data | MS SQL, Oracle SQL, PL/SQL, T-SQL, Hadoop, HIVE, Spark, MapReduce, MongoDB
Analytical Techniques | Hypothesis, A/B Testing, ANOVA, Multivariate Analysis, Regression, Classification, Clustering, Confidence Interval
Cloud Tools | AWS, Azure, Docker, Databricks
Packages | numpy, pandas, scikit-learn, matplotlib, Spark SQL, Spark ML, tensorflow, keras, PyTorch

CERTIFICATION

Azure Data Engineer (DP 200), Tableau Certified Desktop Specialist, AWS Associate Developer, ISTQB Certified Agile Tester

EXPERIENCE

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

- Developing a data science application that predicts real-time health indication of Git repositories and sends periodic alerts.
- Designing an automated data pipeline using **Azure Data Factory** that extracts large training datasets from **Google Bigquery**, pre-processes the data and trains **LSTM** deep learning model on continuous time-series data using **Azure Databricks**.

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

- Automated auditing of support tickets by extracting and analyzing **large unstructured data** using techniques like **NLP**, **Topic Modelling** and **Text mining** in **Python** that increased operational efficiency by 34%, reduced inventory costs by 28%.
- Developed a Time Series Forecasting model in **R** with RMSE of 1.302 to forecast the weekly, monthly, quarterly and yearly demand of network drives usage that reduced server operating costs by 12%.
- Designing data models, interactive dashboards, and ad-hoc reports leveraging data analytics solutions to analyze trends using **Tableau**.

Data Analyst – Robert Bosch, India Nov 2017 – June 2019

- Led a cross-functional team of 3 to enhance data retrieval, manipulation & storage. Aided in training and mentoring of freshers.
- Collaborated with clients, gathered requirements, developed **workflows**, **ETL pipelines** using **SSIS**, importing data from Oracle and Teradata.
- Managed existing data marts, fact & dimension tables, developed queries to fetch results based on user interaction with in-house application.
- Delivered multiple reports in **Power BI** to internal stakeholders and senior management by conducting data mining, data wrangling & exploratory data analysis using **Python**. Helped identify process compliance risks, track ongoing defects saving ~3M dollar revenue for RBEI.
- **Project Management:** Conducted **Scrum** meetings, used **Agile SDLC** methods, conducted code reviews, delivery planning & estimation.

Data Analyst – Verizon Data Services, India July 2016 – Nov 2017

- Created process workflows for internal business customers using **PL/SQL** code, complex **SQL** queries involving triggers, functions, joins & procedures, and integrated them with the **Clarity project management tool** using **XML & gel script**.
- Improved the efficiency of large-scale business processes in Clarity by 21% by **optimizing SQL queries** to enable faster retrieval of records.
- Designed periodic dashboards, ad-hoc custom reports using **SQL**, **Excel & Power BI** to monitor KPIs saving 45 man-hours per week.
- Created an automated **CI/CD pipeline** tool '**DevOps Suite**' using **Java & Jenkins**, which helped in reducing overall lead time by 35%.

DATA SCIENCE APPLICATIONS

ML Application to Predict Healthcare Costs | *Linear Regression, Python, AWS ECR, Fargate, Docker, Flask, Dash, Unicorn, Git*
• Designed a ML application that hosts linear regression model to predict healthcare costs. Deployed application in AWS Fargate for public use.

PROJECTS

- Propensity to Respond Model – Customer Behavior Segmentation** | *R, SMOTE, Random Forest, XGBoost, LASSO Regression*
• Built classification model to predict likely behavior of customers at VMWare by applying Random Forest, XGBoost models with 98.9% recall.
- NLP - Text Mining of Online Physician Reviews** | *Python, R, Gensim, Topic Modelling, Natural Language Processing*
• Extracted reviews, performed feature engineering, identified & quantified effect of key factors on ratings using Linear Regression Algorithm.
- Product Recommendation System** | *Big Data, Collaborative Filtering, Matrix Factorization, PySpark*
• Analyzed 500k reviews & built Recommender System for Amazon fine foods by applying ALS Matrix Factorization model with RMSE 1.204.