Vaishnavii Ponnusamy Paramashivam

Chicago, IL | 312-934-8621 | vaishnavii.pp@gmail.com
linkedin.com/in/vaishnavii-pp/ | vaishnavii28.github.io/ | github.com/vaishnavii28

SUMMARY

- Skilled resource with 4+ years of professional experience in the areas of Data Science, Data Engineering and Data Analytics seeking full time opportunities as a Data Scientist/Data Analyst.
- Proficient in Python, R, SQL, Big Data, Power BI, Tableau, Advanced Statistics, Machine Learning, Deep Learning, 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, Tableau, MS Power BI
SQL/NoSQL/Big Data	| MS SQL, Oracle SQL, PL/SQL, Hadoop, HIVE, Spark, MapReduce, MongoDB
Analytical Techniques	| Hypothesis, A/B Testing, ANOVA, Regression, Classification, Confidence Interval, Sampling
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** predictive model on continuous time-series data using **Azure Databricks**.

Data Scientist – UIC College of Dentistry, Chicago, USA   June 2020 – Present
- Extracted and analyzed **large unstructured data** using **NLP**, **Topic Modelling** and **Text mining** techniques in **Python** that increased operational efficiency and 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 drive usage that reduced server operating costs by 12%.
- Designing & implementing data models, interactive dashboards, and visualizations for ad-hoc reporting in **Tableau**.

Data Analyst – Robert Bosch, India   Nov 2017 – June 2019
- Developed & managed an **ETL data pipeline** that enable flow of vehicle diagnostic data using **SSIS**. Created, modified packages in SSIS as per requirements. Imported large volumes of data from variety of sources, systems including Teradata, Oracle.
- Designed **MLOps** processes and deployed a machine learning application that predicted the criticality of run-time errors in engine software (AUS-ROC 92%), as a web endpoint using **AWS Fargate**, which improved process efficiency by 40%.
- Built multiple reports in **Power BI** by conducting data mining, data wrangling & exploratory data analysis using **Python** that helped identify process compliance risks saving ~3M dollar revenue.
- **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 workflows based on custom business requirements 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**, **Jaspersoft** & **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%.

ACADEMIC 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.

ETL Process – Data Analysis on Live Music Events in UK | *Python, PostgreSQL, Tableau, PySpark, Selenium*
- Designed an ETL pipeline for large unstructured data, ingested transformed data in PostgreSQL data mart, for deriving inferences in Tableau.

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.