PAVANA LAKSHMI VENUGOPAL

pavanalakshmi.v@gmail.com| 518-986-0730 | linkedin.com/in/pavanavenugopal | github.com/datawhiz02

EDUCATION

Master of Science in Computer Science and Engineering (3.73/4 GPA)   December 2023
University at Buffalo, The State University of New York
Relevant Coursework: Pattern Recognition, Analysis of Algorithms, Predictive Analytics, Modern Networking Concepts, Data Intensive Computing, Machine Learning, Data Models & Query Language, Deep Learning on Graphs

Bachelor of Engineering in Electronics and Communication (3.6/4 GPA)   June 2013
Visvesvaraya Technological University

KEY SKILLS

Languages: C, SQL, Python, Spark
Tools: RQM, JIRA, Crucible, Touchstone, Eggplant, Tableau, PowerBI, PostgreSQL, Pytorch, TensorFlow, Microsoft Azure, Git, Linux, Spark, Pandas, Numpy, Scikit, Matplotlib, PowerApps.
Experience in: Version Control Systems, Knowledge on Data Modeling, Data Visualization, Data Ingestion and Extraction.
Soft skills: Outstanding communication / interpersonal skills & ability to work under tight schedules.

WORK EXPERIENCE

Data Engineer Intern, AppMandi LLC, Chicago, Illinois   *June 2023 – Aug 2023*
- Streamlined data extraction, transformation, and loading processes (ETL) using Python scripts, reducing overall processing time by 50% and increased the model efficiency by 40%.
- Developed interactive dashboards in PowerBI to visualize complex data sets, resulting in a 40% increase in data accessibility and understanding for stakeholders.
- Developed a user-friendly app using PowerApps, integrating AI/ML technology to enhance the user experience.

Team Lead| Senior Test Automation Engineer, Cerner Healthcare, India   *Dec 2013 – Jan 2021*
- Managed and coordinated release activities for 3 teams, ensuring timely delivery of all project deliverables.
- Developed and executed comprehensive automated test cases to support software enhancements, resulting in a 95% reduction in manual effort and saving over 500 hours of testing time annually.

ACADEMIC EXPERIENCE

Predictive Analytics: **Python, KNN, Linear & Logistic Regression, Decision Tree, Neural networks**
- Developed and trained predictive models using advanced machine learning algorithms such as random forests, gradient boosting, and neural networks to analyze large datasets.
- Implemented feature engineering techniques such as PCA and variable selection to optimize model performance and reduce dimensionality of datasets by an average of 60%.

Big Data using MapReduce & Spark: **Java, Python, Hadoop, MapReduce, PySpark**
- Developed and executed a scalable data ingestion pipeline using MapReduce and PySpark, resulting in a 50% reduction in processing time for large text datasets.
- Implemented algorithms to analyze and extract meaningful insights from massive text datasets.

Property Tax Analysis in West Roxbury: **Python, Streamlit, Random Forest, Gradient Boosting**
- Created end-to-end data science pipeline, including EDA, data cleaning, model building, and predictions.
- Trained models such as Multiple Linear Regression, Decision Tree, Random Forest Regressor, Support Vector Regression, KNN and Gradient Boosting Regression.
- Developed web-based UI for interactive data analysis and model deployment.

Hospital Information System: **Python, SQL, Database design, PostgreSQL, Flask, HTML, CSS**
- Developed and implemented a dynamic website that seamlessly integrates with our database, enhancing user experience by providing real-time query visualization and results display.
- Boosted data accuracy and compliance in PostgreSQL, modifying storage and retrieval for better patient care.

Photo Editor & Video Compression App: **Python, Django, Computer Vision, reactJS, Microservices, Docker, Kubernetes**
- Delivered a user-centric Passport Photo Maker tool that simplifies image compliance and saves users 1-2 hours and $25 compared to traditional services.
- Devised microservice architecture for 12 features, ensuring modularity, scalability, and independent deployment.
- Orchestrated deployment on Kubernetes cluster using GitHub Actions, automating continuous integration and delivery pipelines.

Credit Card Fraud Detection Models: **Python, Random Forest Classifier, XG Boost, TensorFlow**
- Worked with the Vesta credit dataset, a dataset of 1,097,231 rows and 434 columns, to develop machine learning models.