Yaswanth Pati

+1(862)-230-1268 | yashwanth.pati113@gmail.com | LinkedIn | GitHub | Portfolio

EDUCATION

Saint Peter's University   **Feb 2025**
Master of Science, Data Science
- **GPA:** 3.9/4.0
- **Coursework:** Machine Learning, Deep Learning, Data Warehouse and Business Intelligence, Data Visualization, Applied Statistical Methods, Data Structures and Algorithms, Data Mining

SKILLS

- **Programming Languages**: Java, Python, R, SAS, JavaScript
- **Frameworks**: Node.js, Express JS, Django, Angular, Maven, Selenium
- **Python Libraries/Machine Learning**: OpenCV, Pandas, NumPy, Sci-Kit Learn, NLTK, Matplotlib, Beautiful soup, TensorFlow
- **Database Technologies**: SQL, NoSQL, MongoDB, Oracle, Snowflake, PostgreSQL, Azure Databricks
- **Visualization tools**: Tableau, Looker, Qlik, Power BI, ArcGIS, Excel, D3.JS
- **Cloud/ Big Data Technologies**: AWS, Hadoop (HDFS and Spark), MapReduce, Scala, Azure
- **Big Data Tools**: ETL, Hadoop, Hive, Map reduce, Apache Spark, Data Mining
- **Other**: Docker, GitHub, Rest API, Access, shell scripting, Advanced Microsoft Excel skills, VSCode, Linux, Statistics, Mathematics

WORK EXPERIENCE

Elegantek LLC   **Apr 2025 - Present**
Data Scientist   *Novi, Michigan*
- Developed and deployed ML models using (Random Forest and logistic regression) on historical sales data improving forecasting accuracy by 22% and and enabling optimized inventory planning.
- Designed and implemented dynamic Tableau and Power BI dashboards visualizing KPIs from CRM and financial data, reducing manual reporting by 40 hours/month and accelerating decision cycles for operations teams.

Boven Technologies   **Mar 2023 – Feb 2024**
Data Scientist   *Remote, USA*
- Built predictive models using XGBoost and Random Forest to forecast customer trends with 22% accuracy improvement; deployed Flask-based API on AWS Lambda with CI/CD integration.
- Performed NLP on 100K+ customer reviews using NLTK and SpaCy for sentiment classification and text summarization; insights informed marketing strategy, increasing customer engagement by 15%.

Infosys Limited   **Jan 2021 - Jul 2022**
Data Analyst   *Hyderabad, India*
- Cleaned and transformed operational datasets using Python (Pandas) and SQL, revealing revenue-impacting trends and automating data pipelines with Airflow, which improved efficiency by 30% and reduced processing time by 20 hours/month.
- Designed SQL-powered Tableau and Power BI dashboards tracking KPIs for marketing and finance teams; collaborated in Agile sprints to validate business hypotheses and deliver insights 30% faster.

Lance Soft Engineering Pvt.Ltd   **Jan 2019 - Dec 2020**
Jr. Data Analyst   *Bangalore, India*
- Established and evaluated classification models (e.g., Decision Trees, SVM) to predict equipment failure, improving operational efficiency by 20%; identified root causes and suggested data-driven process improvements.
- Conducted end-to-end data preprocessing and feature engineering on 500K+ records using Python (Pandas, NumPy), resulting in 15 % model accuracy boost on logistic regression models.

PROJECTS

Multi-Model Text Summarization Using NLP   **Feb 2025**
- Developed an NLP-based text summarization system using Transformer models like BART, T5, integrating TF-IDF, BERT Cosine for output evaluation, and built a Flask API with an interactive UI for real-time user input and summary generation.

Route Planning for sensitive populations   **Nov 2024**
- Innovated a mapping tool using Azure Maps API to display live pollutant data and guide users to cleaner routes, with air quality and climate visualizations reducing respiratory risks by 10% and engineered voice-enabled inputs that cut query times from two minutes to under fifty seconds.

Wildfire Prediction Using ML   **Nov 2024**
- Devised a predictive model using MODIS and VIIRS satellite data, achieving over 95% wildfire classification accuracy, and built a confidence-level model leveraging brightness, FRP, and track dimensions, reaching 97% accuracy with a Random Forest algorithm.

CERTIFICATIONS

- **HIPAA Training :** HIPAA Awareness for Healthcare Providers (2025)
- **Databricks Generative AI Fundamentals**: Databricks (2025)
- **Google Data Analytics**: Coursera (Jun 2022)