Sarthak Vajpayee

Dallas, TX, USA | +1(469)-347-9198 | sarthak.vajpayee@utdallas.edu | https://sarthakv7.github.io/my_folio/
https://www.linkedin.com/in/sarthak-vajpayee/

EDUCATION

The University of Texas at Dallas, Texas **Aug 2022 - May 2024**
Master of Science - Business Analytics (Specialization in Data Science)

JSS Academy of Technical Education, India **Aug 2014 - Jun 2018**
Bachelor of Technology - Electronics & Communication Engineering

SKILLS

Programming: Python, R, SQL, NoSQL, MongoDB, JavaScript, C, MATLAB, SAS
Libraries: Scikit-Learn, Seaborn, Spacy, Tensorflow, Keras, OpenCV, PyTorch, Flask, Django, PySpark
Software & Tools: Git, Docker, Kubernetes, AWS, Azure, GCP, Hadoop (Hive, HDFS, Pig, MapReduce), Snowflake, Spark, Scala, Tableau, Excel, Salesforce, NetSuite, Jira, Kafka, ServiceNow, Confluence, CI/CD, Ansible, Grafana, Jenkins, SageMaker
Courses: Applied Machine Learning, Language Processing, Advanced Statistics, Applied Econometrics and Time Series Analysis, Big Data, Predictive & Prescriptive Modeling, Data Warehousing, MLOps, Geospatial Analysis

PROFESSIONAL EXPERIENCE

AppSteer – Frisco, Texas, USA **May 2022 - Dec 2023**
Machine Learning Research Intern

- Led development of **prompt engineering** for **Large Language Models**, enhancing model performance by 30% through collaboration with cross-functional teams utilizing **Langchain, Huggingface** and **PyTorch**.
- Developed **automation** tools with **Flask** and **Python** and **LLM** on **Azure**, integrating **Kubernetes** and **Spark**, reducing **app development** and **deployment** time from 4 days to under 2 minutes.
- Orchestrated over 50 **cloud deployments** and **DevOps automation** using **CI/CD, Ansible,** and **Airflow** on **Azure.**
- Streamlined **API development** with **FastAPI**, delivering a functional interface in six weeks, boosting efficiency by 40%.

Ernst & Young – Bengaluru, India **Dec 2021 - Jul 2022**
Data Scientist

- Refined **predictive modeling,** leveraging **XGBoost and linear models** within Azure, surpassing 90% accuracy and enhancing **demand forecasting** for a leading FMCG company in the Nordic market.
- Engineered a Gen-2 **predictive system**, integrating **ARIMA** and **XGBoost** in **PySpark** over **Hadoop**, achieving 92% accuracy in 15 days and redefining benchmarks for international market analytics.
- Identified critical **demand trends** through **time-series analysis**, resulting in a 15% increase in forecast accuracy, leveraging **PowerBI** and **Vue.js** dashboards for **dynamic visualization**.
- Employed **Azure Databricks** for **data processing** and **Azure ML** for **feature engineering**, enhancing sales forecast models by 20% and yielding $1.1M in annual cost savings.
- Excelled in **dashboard design** exploiting **Power BI**, enhancing decision-making in inventory and sales management with a 15% increase in data retrieval efficiency.

Scaler – Hyderabad, India **Dec 2019 - Nov 2021**
Machine Learning Engineer

- Architected a robust **Learning Management System** and online coding platform using **Docker, Django, and AWS**, integrating an **NLP-based ticketing system** increasing ticket resolution efficiency by 25%.
- Conceptualized and executed a cutting-edge **plagiarism detection** tool via advanced ML models **(Doc2Vec, BERT, Fast-Text)**. This tool has been crucial in maintaining academic integrity across 20k+ code submissions.
- Boosted startup to $50M worth, **doubling revenue** through **tech enhancements and market growth**.

PROJECT EXPERIENCE

- **Research Project**: Led a research project enhancing a **CNN-based (RESNET-50**) COVID-19 diagnostic tool, achieving 0.96 AUC and 92.78% accuracy, rivaling RT-PCR tests.
- **Research Project**: Modeled a **diagnostics framework** for predicting mortality risk of **liver cancer** using MLP classifiers and **Random Forest based MICE imputation**, achieving a 90.02% accuracy.
- **Kaggle**: Employed **Transfer Learning Ensembles** to score in the **top 5%** in Google Q&A Labeling NLP Challenge.
- **Kaggle**: Scored in the **Top 6%** in Mercari Price Suggestion Challenge through l**inear models.**

CERTIFICATIONS

LangChain for LLM App Development | Production Machine Learning Systems | Google Analytics Certification | Deep Learning Specialization | Image Understanding with TensorFlow on GCP | Advanced Machine Learning with TensorFlow on GCP

Eligibility: Authorized for full-time employment in the U.S. for up to 36 months (OPT) without the need for sponsorship.