# Sarthak Vajpayee

Dallas, TX-75252 | (469)-347-9198 | sarthak.vajpayee@utdallas.edu | LinkedIn | Portfolio website

**EDUCATION**

**The University of Texas at Dallas,** Texas 05/2024
*Master of Science, Business Analytics (specialization in Data Science)* **GPA 3.80**
Courses: Applied Machine Learning, Advance Statistics, Advanced Deep Learning, Applied Econometrics, Advanced Natural Language Processing, Database Management.
Academic awards: **Dean's Excellence Scholarship** (*Jindal School of Management*)

**JSS Academy of Technical Education**, Noida, India 06/2018
*Bachelor of Technology, Electronics & Communication Engineering*

**SKILLS**

**Certifications:** **Google:** Advanced Machine Learning with TensorFlow, Deep Learning Specialization
**Stanford University:** Machine learning, Natural Language Processing with Python
**Programming:** Python, R, MySQL, MongoDB, JavaScript, C, MATLAB
**Libraries:** Scikit-Learn, Seaborn, Spacy, Spark, Tensorflow, Keras, OpenCV, PyTorch, Flask, Django
**Software:** Git, Docker, Kubernetes, AWS, Azure, GCP, Hadoop, Scala, Hive, Tableau, Power BI

**PROFESSIONAL EXPERIENCE**

**AppSteer –** Frisco, Texas, USA 05/2023 – 12/2023
*Machine Learning Research Intern*

- Initiated and led the development of **prompt engineering** for **Large Language Models**, enhancing model performance by 30% through collaboration with cross-functional teams using **Langchain** and **Huggingface**.
- Streamlined **API development** using **FastAPI**, delivering a fully functional interface in just six weeks, which boosted company's operational efficiency by 40%.
- Developed an **automation** tool 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.

**Ernst & Young Global Limited –** Bengaluru, India 12/2021 - 07/2022
*Data Scientist*

- Leveraged **XGBoost and linear models** within Azure to refine **predictive modeling**, surpassing 90% accuracy, significantly enhancing **demand forecasting** for a leading FMCG company in the Nordic market.
- Rapidly 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.
- Utilized Python's statistical libraries and **PyTorch** for comprehensive **time-series analysis**, revealing critical demand trends that were dynamically **visualized** with **PowerBI** and **Vue.js** dashboards.

**Applied Roots –** Hyderabad, India 09/2019 - 11/2021
*Machine Learning Engineer*

- Developed a robust **LMS** and online coding platform using **Docker, Django, and AWS**, integrating an **NLP-based ticketing system** that increased ticket resolution efficiency by 25%.
- 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**: Developed 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**: Mercari **Price Suggestion Challenge**, scored in the ***Top 6%*** *using only* ***linear models.***
- **Kaggle**: Google Q&A Labeling **NLP** Challenge, ***Top 5%*** *using* ***Transfer Learning Ensembles***.