Vatsal Vinay Parikh

parikhvatsal7101@gmail.com | +1(930) 215-5267
Bloomington, IN, 47401| LinkedIn | GitHub

Education:

Indiana University, Bloomington, IN, United States — *August, 2023 – Present*
Masters of Science in Data Science — *GPA: 3.83 / 4.0*

Gujarat Technological University (GTU), Ahmedabad, India — *July, 2019 – June, 2023*
Bachelors in Engineering (B.E.), Information Technology — *GPA: 9.3 / 10.0*

Coursework:

Data Mining, Applied Machine Learning, Introduction to Statistics, Usable Artificial Intelligence, Applied Database Technologies, Social Media Mining, Stanford University – Machine Learning Specialization, IBM Data Science Professional Certificate, Architecting with Google Compute Engine Specialization, Amazon Web Services (AWS) Fundamentals, Google IT Support Professional Certificate, Intermediate R, UC San Diego – Big Data Specialization

Skills and Expertise:

- **Programming Languages:** Python, SQL, R
- **Libraries:** Pandas, NumPy, Matplotlib, OpenCV, SciKit-learn, NLTK, TensorFlow, Keras, PyTorch, ggplot2
- **Database Management**: MySQL, PostgreSQL, MongoDB, Neo4j, Cassandra, Snowflake
- **Other Tools:** AWS, Google Cloud, Linux, Git & GitHub, PowerBI, Tableau, Kubernetes, Docker, Apache Hadoop

Work Experience:

Data Analyst Intern, CSRBOX Foundation, Ahmedabad, India

- Worked on developing a hands-on machine learning project **"Heart Disease Prediction System"** under the guidance of mentors from IBM.
- Used Python libraries such as **sci-kit learn, streamlit** to develop an interactive web application that predicts the presence of a heart disease based on several medical parameters.
- Applied **Support Vector Machine (SVM)** to achieve the highest **accuracy of 93%** for classification.

Research Assistant, Kelley School of Business, Indiana University, Bloomington

- Worked as a Faculty Assistance in Data Science student to employ AI and machine learning techniques over **100,000 on-line conversations** related to mental health by extracting data from **Reddit API** using **PRAW**.
- Orchestrated the **cleaning and analysis** of panel data from a leading weight management platform and applied **advanced machine learning** algorithms to identify distinct user engagement patterns.
- Applied **NLP algorithms** in Python to accurately find personality traits from unstructured social media data and integrated **clustering techniques** to classify users into distinct **personality clusters**.

Academic Projects:

Predictive Modelling for Personalized Diabetes Care

- Spearheaded data mining techniques such as data cleaning, pre-processing, decision-making, and correlation analysis on a comprehensive dataset comprising **70,692 subjects** from the Behavioral Risk Factor Surveillance System (BRFSS).
- Developed machine learning models achieving up to **75.47% accuracy** in predicting diabetes risk using **Artificial Neural Networks (ANN)** and created a web-based application for real-time risk assessment through streamlit.

Application Failure Prediction Model

- Created a machine learning model that predicts the failure of application executions on Purdue ITAP's Central Computing cluster dataset.
- Achieved highest balanced accuracy using **DecisionTreeClassifier** and **FastICA** machine learning algorithms along with **Principal Component Analysis** for feature selection.

Waste Management and Garbage Classification Model

- Developed an **AI-powered waste management system** as part of the Luddy Hackathon securing rank 4, by utilizing **InceptionV3 transfer learning** and fine-tuning techniques, achieving accuracy rates of **89.37%** and significantly improving waste sorting efficiency and accuracy.
- Integrated **real-time video processing** capabilities into a user-friendly web application using the **Streamlit** framework, enabling efficient waste classification and providing disposal guidance to users.

Research Papers:

- Researched and authored a paper titled '**Analyzing Twitter Reactions to the Tesla Cybertruck: Before and After Launch**,' which involved conducting content analysis on Twitter discourse surrounding the Tesla Cybertruck, **evaluating shifts in public sentiment** and engagement metrics pre- and post-launch, and providing insights into consumer perceptions of automotive products.
- Researched and authored a paper titled '**Analysing the Public Discourse around OpenAI's Text-To-Video Model 'Sora' using Topic Modeling'** which involved conducting topic modeling analysis on Reddit data, identifying dominant themes and topics, and visualizing insights using natural language processing techniques.
- Researched and authored a paper titled **'Exploring the Non-Fungible Token (NFT) Discourse: A Sentiment Analysis of User Discussions on Reddit'** and conducted sentiment analysis on Reddit discussions surrounding NFTs revealing shifts from initial positivity to growing skepticism amidst market fluctuation.