Kaustubh Vora

Denton, TX 76201 | (940) 629-6795
Kaustubhvora60@gmail.com | https://www.linkedin.com/in/kaustubh-vora-887752170/

EDUCATION

University of North Texas - Cumulative GPA 4.00 08/2023 - 05/2025
*Master of Science, Major in **Data Science***

NMIMS UNIVERSITY - Cumulative GPA 3.56 07/2019 - 05/2023
*Bachelor of Technology, Major in **Computer Science***

PROFESSIONAL EXPERIENCE

Data Analyst Intern - *NTT Data Payment Service* 05/2022 - 11/2022

- Specialized in data visualization and **Data pre-processing,** developed data processing scripts, resulting in a 20% reduction in data cleaning time.
- Performed various feature engineering techniques such as OneHotEncoding, Frequency Encoding, Ordinal Encoding.
- Conducted data interpretation, identifying patterns and anomalies to provide insights for security strategies.
- Assisted in the development and implementation of machine learning models to forecast merchant's demand.
- Actively participated in process improvement initiatives, streamlining data analysis workflows by 15%.
- Collaborated with the team to prepare comprehensive statistical reports and presentations summarizing key metrics and trends with effective data visualizations.

Data Analyst Intern - *CleverCheckIn* 12/2020 – 05/2021

- Successfully managed and analyzed travel data streams, including booking patterns, seasonal fluctuation, and market trends.
- Utilized advanced time series analysis to forecast booking patterns leading to 13% improvement in accuracy compared to previous methods.
- Analyzed historical booking data to identify revenue optimization with **ARIMA** model, exponential smoothing.
- Worked closely with marketing, sales, and product development team to understand business objectives, translate them into actionable analytics tasks.
- Created interactive dashboards using **Tableau**, enabling stakeholders to intuitively explore trends and make decisions, resulting in a 8% increase in decision making process.

PROJECT EXPERIENCE

Movie Recommendation Engine

- Spearheaded the development of a sentiment analysis system tailored for IMDb movie reviews, employing advanced **NLP** techniques.
- Successfully addressed challenges such as stop words removal, English Contractions and lemmatization, ensuring the datasets readiness for subsequent machine learning tasks. Employed count vectorization and TF-IDF vectorization to enhance the representation of lemmatized tokens, contributing to 11% improved model performance.
- Evaluated and implemented a diverse set of models tailored for sentiment analysis, including Logistic Regression, Decision Tree Classifier and Random Forest Classifier.

Image In-Painting & Colorization

- Developed a model, where a damaged image is in-painted and colorized with the machine learning algorithm.
- Led the implementation of a robust Generative Adversarial Network (GAN) architecture to tackle image restoration challenges, effectively reconstructing damaged regions in images.
- Engineered a customized solution that learned intricate color patterns and relationship within images, contributing to the revitalization of visual content.

Stock Market Analysis

- Utilized Stock market KPIs to determine the state of the tech sector of the stock market.
- Maintained a working knowledge of data ETL, Visualization, API frameworks and statistical methods in **Python**.
- Cleaned Data and accessed JSON endpoints for 15+ companies.
- Conducted a regression to determine the correlation between NASDAQ's price to earnings ratio to revenue growth.

CERTIFICATES

- Data Science & Machine Learning Bootcamp (Udemy), Tableau (Udemy)
- MYSQL (Hacker Rank), Microsoft Excel (LinkedIn Learning), Data Analytics (Coursera)
- The Business Intelligence Analyst Course (Udemy)

SKILLS

- **Analytical techniques:** Pandas, Numpy, Probability, Statistics, ARIMA, NLP, Data Manipulation, Regression, GCP, AWS.
- **Analytics platforms:** Python, R Programming, Command Line, Excel, JAVA, Git and Version Control, SPSS, MySQL.
- **Visualization Platforms:** Tableau, PowerBI, Python.
- **Deep Learning Techniques:** TensorFlow, PyTorch.
- **Languages:** English, Hindi, Gujarati.