MOHIT KATTA

mohit.kvs1@gmail.com • (971) 348-1275 • Boston, MA • https://linkedin.com/in/mohit-katta

EDUCATION

Boston University| Boston, Massachusetts
Masters of Science in Applied Data Analytics Sep 2022 - Dec 2023
GPA: 3.8 / 4.0.
Relevant courses: Advanced Machine Learning, Data Mining, Big Data Analytics, Data Science with Python.

Heriot-Watt University| Dubai, UAE
Bachelors of Science (Hons.) in Computer Science Sep 2019 - Jun 2022
GPA: 4.0 / 4.0.
Relevant courses: Introduction to AI, Data Mining and Machine Learning, Computer Network Security, Software Engineering.

EXPERIENCE

Boston University Henry M. Goldman School of Dental Medicine| Boston, MA Dec 2022 - Dec 2023
Graduate Student Data Science Researcher

- Examined 40,000+ American's tweets related to Vaccine and fluoride, to identify trends in beliefs.
- Reduced manual assigning of sentiments on tweets by 40%; created a ML pipeline to perform Pre-processing, fine-tuning a transformer model to automatically assign sentiments.
- Fine-Tuned an existing Transformer model to predict the sentiment of more than 40,000 tweets with an accuracy of 85%.
- Developed correlation matrices and Bar plots to assess the trends over 3 time periods: Pre Pandemic, During Pandemic and post pandemic.

BU Metropolitan College IT| Boston, MA Sep 2022 - Dec 2023
Graduate IT technician

- Resolved 200+ tickets within a span of 1 year with over 90% client satisfaction rate.
- Troubleshooted hardware and software issues such as Power BI, VMWare and Citrix Virtual Labs via ServiceNow ticketing system, phone calls and Bomgar remote desktop assistant.
- Trained a team of 4 freshers to assist clients with various troubleshooting tasks such as Citrix Virtual lab issues, printer issues and handling clients face-to-face over a period of 3 weeks.

Apparel group| Dubai, UAE Jun 2022 - Jul 2022
Data Science Intern

- Optimized the inventory for the Charles & Keith brand by developing a sales predictive algorithm; Reduced unused inventory by an average of 15% per store.
- Analyzed diverse data factors, such as customer demographics, store locations, seasonal trends, store sizes, and fashion cycles, to enhance the algorithm's performance.
- Deployed a Sales Predictor Model, with an accuracy of 75% at the time of deployment, providing valuable insights and optimization solutions for the brand.

ACHIEVEMENTS

Deputy Principal's Award - Heriot-Watt University Sep 2021 - Sep 2022

- Received award for Exemplary performance in Academic activities; 4.0 GPA

CBSE Topper's Award - High School May 2019

- Received award for outstanding grades in Senior Secondary level.

PROJECTS

Trends in American's Beliefs about Fluoride from Twitter Dec 2022 - Dec 2023

- Conducted in-depth analysis of public sentiment regarding water fluoridation. Employing advanced techniques to extract 80,000 relevant tweets leveraging a Web Scraping tool via Digital Ocean VPC.
- Collected a subset of 1000 tweets for manual labeling; Later used for fine tuning a Transformer model (RoBERTa).
- Created a ML Pipeline to pre process the data, normalize and remove unnecessary tweets for fine-tuning the transformer model.
- Analyzed the sentiment of 40,000+ tweets with the Fine-Tuned RoBERTa model, enabling effective predictions on a large volume of unlabeled data.
- Developed correlation matrices by clustering topics derived from unstructured data sources.

Stock price prediction using past stock prices and tweets. Sep 2022 - Dec 2022

- Collected more than 5 years worth of stoick prices for the Alphabet stock ticker. Conducted predictive analysis on historical stock prices and public sentiment data to project future stock prices.
- Explored and evaluated the performance of 9 different Machine Learning models and three word embedding techniques, namely TF-IDF and "Bag of Words" for analysis.
- Implemented a data pipeline for daily Twitter data processing, normalization, and sentiment analysis.
- Identified "Bag of Words" as the optimal word embedding model and Support Vector Classifier (SVC) as the most effective classifier.

Ensemble Machine Learning Model for sentiment analysis. Sep 2021 - Mar 2022

- Developed an ensemble Machine Learning model by combining two Deep Leaning models with TensorFlow and OpenAI; Utilized SpaCy and NLTK libraries to prepare the extracted data for seamless model training.
- Utilized Digital Ocean VPCs to establish connections with multiple data centers worldwide, allowing for the simultaneous extraction of terabytes of tweets for data acquisition.
- Orchestrated a robust data pipeline utilizing SpaCy and NLTK libraries to preprocess and prepare the extracted data for seamless model training.
- Evaluated the model with Accuracy, MCC Coefficient and other ML metrics to reveal an impressive accuracy of 77% for ensemble model, while the existing state-of-the-art model achieved approximately 85%.

Data Mining and Machine Learning - Portfolio. Sep 2021 - Mar 2022

- Implemented various NLP techniques such as Naive Bayes, k-means clustering, hierarchical clustering, decision trees, and linear classifiers, all evaluated using a 10-fold cross-validation approach.
- Employed a comprehensive set of evaluation metrics, including Accuracy, F-Score, ROC/AUC curve, precision, and recall, to assess the effectiveness of each model.

TECHNICAL SKILLS

- Python, R, Dart, Ocaml, SML, Prolog, LaTeX.
- TensorFlow, HuggingFace, PySpark, Pandas, NumPy, Sci-Kit Learn, Seaborn, React JS.
- Amazon Web Services (AWS), Google Cloud (GCP), Apache Spark, Apache Kafka.