VARUN KABRA

San Jose, CA (Open to Relocation) | +1 4086906426 | kabra.va@northeastern.edu | LinkedIn | Portfolio | GitHub

EDUCATION

Northeastern University *Sep 2022 - May 2024 (Expected)*
Master of Science in **Data Science**, GPA: **3.85/4.0**
Related Coursework: *Deep Learning, Generative AI, Supervised, Unsupervised Machine Learning, Natural Language Processing*

The University of Texas at Austin [e-Portfolio] *Feb 2021 - Mar 2022*
Post Graduate Program in **Data Science & Business Analytics**

Manipal Institute of Technology *Jul 2017 - Aug 2021*
Bachelor of Technology in **Information Technology**

TECHNICAL SKILLS

Programming Languages: Python, Perl, R, Java, C++
Machine Learning Libraries: Pandas, NumPy, Seaborn, SciPy, Sklearn, Keras, TensorFlow, PyTorch, Matplotlib, Plotly
Cloud Technologies: Azure, AWS: S3, SNS, RDS, Lambda, DynamoDB, EC2, Google AI
DBMS: SQL, MS Access, SQL Server, Oracle Database, PostgreSQL
Other: PowerBI, Tableau, jQuery, Hadoop, Hive, Apache Spark, Excel Macros, Shell Scripting, ELK Stack

WORK EXPERIENCE

Apexon, Data Engineer Intern *Jul 2023 - Dec 2023*

- Transformed data for issue classification analysis by implementing pipeline to transfer Apexon's issue request data to **Azure** for storage, followed by exploratory data analysis in **Databricks** using **Python**.
- Achieved significant reduction in response time for employee-raised issues by increasing issue labeling efficiency by **25%**, employing **Naive Bayes** and **Random Forest** classification models coupled with **NLP** libraries: **spaCy, nltk, scikit-learn**.

Unilever, Data Engineer *Jun 2021 - Aug 2022*

- Accomplished **20%** increase in on-time delivery performance of vendor partners by launching collaborative manufacturing platform, orchestrating seamless communication between Unilever's Supply Chain team and 3rd party vendors.
- Obtained **79%** accuracy in forecasting future costs for ocean freight shipments by leveraging historical shipment quotations, performing feature engineering, extracting vital features like vessel type, volume, source, destination, price.
- Deployed time series forecasting models: **ARIMA, SARIMA**, **LSTM** to predict shipment costs, enabling Ocean Logistics team at Unilever to prepare accurate quotations for shipment negotiations with vendors, saving over **$1M** annually.
- Established two data migration pipelines using **Apache Airflow:** The first moved **7TB** data from AWS S3 to Azure blob storage, followed by transferring to the local Unilever directory, resulting in annual cost savings of **10%.**

Honeywell, Data Engineer Intern *Jan 2021 - Jun 2021*

- Orchestrated creation of return-to-workplace portal for **100,000+** Honeywell employees utilizing **Tableau**, **PowerApps**, **SharePoint**, ensuring seamless transition following COVID-19 restrictions and compliance with safety guidelines.
- Led a team of **5** to build an **unsupervised machine learning anomaly detection algorithm** (**81%** accuracy), enhancing Honeywell's preparedness against viruses by predicting downtime of **40,000+** systems using live, historical system data.
- Attained an **80%** reduction in data latency and improved data accuracy by automating real-time updates between **SharePoint** and **ServiceNow** using **Microsoft SSMS** and implementing button-driven automation through **SQL** queries.
- Crafted self-help employee portal using **Microsoft PowerApps and Python**, pivotal during COVID-19 restrictions, **optimizing resource allocation**, and ensuring business continuity by enabling **remote issue resolution**.

TECHNICAL PROJECTS

Northeastern University, EmoTune: Mood-Enhanced Spotify Playlist [GitHub Link] *Jan 2023 - Apr 2023*

- Developed sentiment analysis technique using **CNNs and RNNs** on a **15,000**-record dataset, achieving over **90%** accuracy in predicting user emotions, while programming end-to-end solution with **LightGBM** for music attribute clustering.
- Employed **Python, NLTK, Scikit-learn, TensorFlow, Keras** for sentiment analysis, integrating **Spotify Web API** for music clustering, achieving **95%** accuracy in clustering **top 50 songs** based on emotional attributes, enhancing user mood.

Northeastern University, MRI Image Categorization [GitHub Link] *Jan 2023 - Apr 2023*

- Facilitated the development of MRI image categorization system using **ResNet50 CNN**, **SVM**, and **hierarchical clustering**, achieving **96%** accuracy in classifying 50000 brain MRI images.
- Implemented **preprocessing** techniques like **Gaussian Blur** and **Principal Component Analysis** extracting **51** features capturing **98%** variance, addressing challenges of class imbalance and variable image clarity.
- Introduced robust data augmentation, significantly enhancing the training dataset for improved generalization performance.

ACHIEVEMENTS

- Achieved **3rd** place in Brembo Hackathon by leading the development of a **generative AI** model (**97%** accuracy) for synthetic data generation enhancing warranty classification accuracy and efficiency. [GitHub Link]
- Winner of **Economic Times Campus Stars 4.0**: India's largest hunt for brightest engineering minds. [Link]