Manish Maurya

DOB: 11/09/1994

GitHub : https://github.com/ManishMaurya11

Contact: +91 9807488789

LinkedIn : https://www.linkedin.com/in/manish-maurya-6596121b3

Email: *manishmaurya1193@gmail.com*

ACADEMIC QUALIFICATIONS

Post-Graduation:	Master of Technology (Aerospace Engg.)	**IIT Bombay**	**2021**
Graduation:	Bachelor of Technology (Mechanical Engg.)	**GNIOT,Gr.Noida**	**2016**

DATA SCIENCE & MACHINE LEARNING COURSES

Engineering Data Mining and Applications | *Guide: Asim Tiwari* | IIT Bombay *(Aug'20-Dec'20)*

- Data mining methods and Applications, Machine learning algorithms like **Clustering**, Classification, **Regression**, Optimization algorithms like Gradient Descent. Recommendation engine, **Big Data** Processing and Analytics, Natural Language processing, LLM, NER, RNN, NLTK, TensorFlow, Keras, Spacy, Transformers (BERT, GPT)

Statistics for Decision Making| *Guide: Usha Ananthakumar* | IIT Bombay *(April'23-May'23)*

- PDFs, Bayes theorem, Distributions, Central limit theorem, Hypothesis Testing, Z-score, T-score, P-value, Analysis of Variance (ANOVA Test)

WORK EXPERIENCE

FlairMinds Software Pvt. Ltd. Pune (Data Scientist) *{May'22-Feb'23}*

Project 1 Churn Rate Analysis and Recommendation Engine Development | Client Name: PayPoint India Network Pvt. Ltd.

- Business understanding of the **finance** company for retailers through Statistical Analysis (**EDA**) of transaction data. Used python (NumPy, Pandas and Seaborn) to analyze and visualize the data. Utilized various **Feature Engineering** for outlier, binning, log transform, normalization.
- Product Level **Time-Series analysis** of amount generated per Business on daily/weekly/monthly basis. Which leads to the Ranking of the Business personals according to the risk of churn. Represented the analysis on **Google Data Studio**.
- Discussed about the probable causes of churn and the corrective measures to take in future to reduce churn rate.

Project 2 Data Analysis using SQL | Client Name: Greenfinch

- Extracted the data from client's Mysql server and transformed it into useful tables using **SQL.** Analyzed the data based on different disease and made patient journey in the hospital. Presented the visualization on **Looker Studio** to doctors and stakeholders.

Project 3 Product Similarity Matching between Products from Amazon, Walmart etc. | Client Name: Iksula Services Pvt. Ltd.

- Modelling for similarity matching of approx. 2.6 million product data. We used **Natural Language Processing** using **LLM** (e.g. **Sentence-transformer/all-mpnet-base-v2** of **HuggingFace**) for text matching between levels of product information such as manufacturing brand, product title, description and attributes. Calculated similarity score between products using **Cosine similarity** on **Word2Vec** embedding. For better speed integrated **Faiss index** for similarity score. Improved model by treating alphanumeric and numeric part separately (using **RegEx**).
- Later improved matching using **Computer Vision** Model **Inception V3** for product image matching.
- **Vertex AI** and **BigQuerry** of **GCP (Google Cloud Platform)** was used for modelling and querying the dataset and output score.

Project 4 Deployment and functionality addition for the NLP based document search solution | Client Name: OnPepper

- Deployment of the solution developed by HummingBirdAI for searching within a large pool of financial documents.
- **Docker** containers for the frontend (**Flask**), backend, database (MySQL) and **Elastic search** server. Used Docker's concepts of Networking, Volumes and Docker-Compose. Developed working understanding of Elastic Search, Logstash and Kibana.

Academic and Personal Projects

Project 1 Movie recommendation System Academic Project| IIT Bombay (Asim Tiwari)

- Build recommender engine for the Kaggle dataset of Netflix and TMDB dataset. Applied Demographic filtering, Content Based filtering and Collaborative filtering to recommend the movies based on popularity as well as custom metric. Harnessed the user and movie ratings as well as characteristics to recommend and rate the unrated movies.
- Utilized **Scikit-learn** library for **vectorization** (TfidfVectorizer) of text and **cosine similarity** as a similarity metric.
- **Surprise** library along with **SVD** (singular value decomposition) algorithm was used for collaborative filtering.
- In addition to the solution a business prospect consisting of product advertisement video and product pamphlet was prepared.

Project 2 Prediction of heart disease | Self (Kaggle)

- Collected the dataset from Kaggle and applied **Data cleaning, feature reduction, One-Hot encoding** and then applied **Decision Tree** and **XGBoost** models, used various metrics to evaluate model like accuracy, confusion metrics, precision, recall and F1 score. **Hyperparameter tuning** was done using **GridSearchCV** for each model to get the higher recall score.

Project 3 Sentiment Analysis | Self (Kaggle)

- Performed sentiment analysis on twitter data for positive and negative classes using **LSTM**. Analyzed the effect of number of tokens during training, effect of cleaning the data, effect of multiple dense layers and the effect of pretrained word vectors using word2vec.
- Negation of the sentence using words like 'not' was successfully captured by neural network.

Project 4 Weather App | Self (Kaggle)

- Made weather app using weather API service, Utilized Flask and basic html to build basic form to take POST request from user and Postman.

POSITION OF RESPONSIBILITIES

Managed projects and built team at Flairminds Pvt. Ltd. *(May'22-Feb'23)*

- **Conducted interviews**, built a capable team of data engineers and data scientists from scratch, **managed** the ongoing projects.

Teaching Assistant (TA) |*Flow Over Circular Cylinder* | IIT Bombay *(Jan'20-June'20)*

- Mentored **50+** B. Tech students for performing experiments using sensors and measurement devices in the Aerodynamics lab.

TECHNICAL SKILLS	Python, OOPS, Pandas, TensorFlow, Keras, Scikit-learn, Spacy, Nltk, GCP (Cloud) SQL, Google Data Studio, MS Excel, MS Word, MS PowerPoint RESTful API, Flask, Django Git, Docker
SOFT SKILLS	Research and brainstorming, Self-management, Teamwork, Conflict Resolution and Adaptability