Ravi Ranjan

Data Scientist

+91 9504189028 | Gaya, Bihar

raviranjan0631@gmail.com | https://www.linkedin.com/in/raviranjan0631/

SUMMARY

I am an Engineer who found his passion in statistical modeling and Algorithm development. I am interested in Artificial Intelligence (AI) / Machine Learning (ML) and the potential they have to offer in solving real-world problems. I am looking forward to working in an environment where I get the opportunity to turn data into products with actionable insights.

WORK EXPERIENCE

HRS Group, Remote

Data Scientist   **APRIL 2021 – Present**

Dynamic Hotel Price

- Forged a regression model using **XGBoost** that predicts the price of the Hotels with respect to destination and amenities with **MAPE of 7% only**
- Created the interactive dashboard using **Microstrategy** for the client to use, saved the **overall time by 93%**
- **Deployed using AWS Apigateway, Lambda** and Sagemaker instance to get a latency of 0.4 second

Mission Control

- Developed a serverless application to recommend similar JIRA tickets with **BERT sentence transformers** and **AWS lambda** that led to a reduction of service time of Jira tickets by **83%**
- Used **quadrant vector DB** to store the embeddings and deployed it on **EC2** instance
- Crafted various **data pipelines** and linked them to interactive Dossier in **Microstrategy** to get the overall report in a single place it reduced the **overall time by 97%**

Hotel Amenities Rating

- Used **NLP** and **aspect-based sentiment analysis** to rank the hotel amenities using reviews from email and customer's comment
- Deployed the application using AWS Lambda and AWS

No Show Prediction

- Developed an end-to-end product **Decision Tree** which predicts whether a booking is fake or not and decreased the time by **84%** and cost by **70%**
- Deployed the model on **Apache Airflow** using **AWS Glue** to update the dashboard on a daily basis

EMAIL TYPE CLASSIFICATION

- Created an email type segregator using **BERT NLP** to filter out the emails and forward them to the correct team, it decreased the overall response time of the email by **60%**
- Deployed the model on **Apache Airflow** to check and forward the emails on an hourly basis

PII Data Anonymization
- Used serverless approach to mask the sensitive data within the data pipeline AWS **AWS Kinesis Firehose**
- Used **DynamoDB** and **AWS Lambda** for the vault mechanism which led to a reduction of overall time by 60%

Rehearse Labs, New Delhi
Data Scientist   **Sept 2020 - Aug 2021**
Auditions
- Used **face recognition one shot learning** to detect the cheating during the live test
- Create the **data streaming using AWS kinesis** to store the keystroke made by the candidate to detect cheating and recreation of candidate test
- Used Miscrostrategy to do the reporting of the candidate over all performance and filtration
- Used **Transformer** to detect plagiarism between code.

Ank Aha Private Limited, New Delhi
Data Scientist   **Jan 2020 – Sept 2020**
Jaano India
- Decrease the cost of Google API to find the total number of schools in India using the **Clustering** approach, decreased the overall cost by **80%**
- Used **multiprocessing** in **python** to decrease the overall API call time by **16 times**
- Removed the abusive posts on the Jaano India apps using **Spacy** for text and **CNN** classifier to remove the NSFW images

RMgx, Gurgaon
Machine Learning Intern   **Jul 2019 – Oct 2019**
GIS Segmentation
- Engineered GIS segmentation of buildings, roads & lakes using robosat.pink package and utilized sliding window segmentation to generate newly labeled data
- Implemented **CNN** from a research paper for remote sensing and achieved similar results with an error of 0.1%.

Projects
- **3D Print Estimator**: performed data cleaning to remove outliers and feature engineering on the data, trained a **Random Forest Regressor** on data, and decreased **RMSE** from **81 to 27**
- **Amazon Fine Food Reviews:** used **tSNE** for the visualization, AUC curve as the KPI, Used **Linear SVM** and **Random Forest**, **Naive Bayes**, and **Logistic Regression** for classification to improve the AUC score
- **Mind Map System**: used **LLM** and **chat gpt4** to create a mind map based on prompts
- **Seismic Detection**: developed a deep learning solution using U-NET CNN architecture for segmentation of salt from seismic image to detect the salt in the image

EDUCATION
Plaksha Tech Leader Fellowship, Mohali   **2021**

PG in AI, **CGPA –** 3.46/4
Greater Noida College of Technology, Noida   **2018**
Bachelor of Technology in Computer Science and Engineering, Aggregate Percent - 76%

SKILLS
Programming Language and Frameworks
Python, Scikit-learn, Hugging Face, PyTorch, TensorFlow, Keras, SQl, Pandas, NumPy, OpenCV, Apache Airflow, Beautiful Soup, Selenium, RDBMS, NoSQL, Docker, Kubernetes, Github, Microstrategy, Qdrant
Deep Learning (DL)
Computer Vision, AutoEncoders, LSTM, RNN, Transformer, NLP, CNN, OCR, Object Detection, Image Processing, Recommender systems
Statistical Modelling and Machine Learning (ML) Techniques
Linear Regression, Kmeans, Random Forest (RF), Decision Tree, Naive Bayes, KNN, Logistic Regression, Collaborative Filtering, Recommendation Engine, Time Series Forecasting, Data Mining, ETL, EDA, Analytics, Visualizations, XGBoost
Cloud Development
AWS (Lambda, Firehose, Athena, Data Lake), Apache Airflow