ANUJ BAGHEL

Data scientist

📞 9004672502   @ anujbagh300@gmail.com
🔗 Linkedin : https://www.linkedin.com/in/anuj-baghel-441518191/
🔗 Github :   https://github.com/bagh2/Anuj-Baghel   📍 Mumbai, Maharastra
☆ Medium :https://medium.com/@anujbagh300

KEY-SKILLS

Python · MYSQL · DATA ANALYSIS ·
DATA MINING · PROBABILITY ·
STATISTICS · MACHINE LEARNING · NLP ·
NAURAL NETWORK ·
RECCOMENDATION SYSTEM ·
POWER-BI · AWS-EC2 ·
MICROSOF AZURE ·
TENSORFLOW & KERAS · SK-LEARN

SUMMARY

A passionate data scientist having knowledge in predictive modeling, data processing, anML algorithms to solve challenging business problems. Good background in Python andknowledge of various types of machine learning , Deep learning , Nlp techniques.

CURRENT PROJECT

CUSTOM CHATBOT ON INDIAN CONSTITUTION DATA

03/2023 - Ongoing   Mumbai, Maharastra
- Developing the custom chatbot on indian constitution data and laws and regulatory data will deploy using **openai-api**

PROJECTS

Healthcare provider claim fraud detection : ML
01/2023 - 02/2023
- Conducted healthcare claim provider fraud detection project with Kaggle data, including 4 CSV files.
- Encountered highly imbalanced data (only 9% of providers were fraudulent). Used SMOTE and ADASYN to handle the imbalance problem.
- Tried various neural network models, but settled on an ensemble model with LR, RF, XGBoost, and SVM as base learners and an 80:20 ratio of SMOTE oversampling.
- Achieved a 77% F1 score, good precision and recall, and a 93% AUC-ROC score with the ensemble model.
- Used F1 score, AUC-ROC score, and confusion matrix as performance metrics.
- **python , Sampling techniqes: smote , adasyn , ml algorithms , ensemble techniques , neural networks, aws. https://github.com/bagh2/Healthcare-claim-provider-fraud-detedtion-**

Youtube video summerizer : NLP
02/2023 - 03/2023
- Developed a YouTube video summarizer using a transformer-based pre-trained model from the Hugging Face library.
- Utilized the YouTube transcript library to extract video text and summarize a 40-minute speech by the Indian Prime Minister to Interpol, with plans to expand to a web application.
- **modules : youtube transcript, hugging face , transformer, bert**
- **https://github.com/bagh2/Youtube-video-text-summerizer-**

EDUCATION

B:Sc: Information technology

Mumbai university

2019 - 2022   Mumbai

CGPA | **9.3** / 10

ACHIEVEMENTS

- 💎 Successfully solved the Microsoft Malware Detection problem using a 250GB dataset on Colab. Demonstrated strong skills in data analysis and machine learning.

- 📑 26+ GITHUB REPOSITORIES

Facebook Friend reccomandation : ML
11/2022 - 12/2022
- Successfully completed a Facebook friend recommendation project using Kaggle data provided by Facebook for research purposes.
- Implemented various techniques to measure distance, including rank indexing, Jaccard distance, cosine similarity, Adar indexing, computed the Katz Centrality and Hits score, and added features such as num followers and num followee.
- Applied SVD factorization and preferential attachment techniques to add new features for followers and followees.
- Achieved high performance in both training and testing by training the model with different models, including Random Forest and XGBoost, and performing feature selection to identify the most important feature for friend recommendation, namely "follow back". The Random Forest model achieved a test f1-score of 0.926, while the XGBoost model achieved a test f1-score of 0.927.
- **Techniques utilized include rank indexing, Jaccard distance, cosine similarity, Adar indexing, Katz Centrality, Hits score, SVD factorization, preferential attachment, Random Forest, and XGBoost.**
- **https://github.com/bagh2/Facebook-Friend-Recommendation**

Microsoft Malware detection : ML
Date period
- Microsoft malware detection problem Successfully reduced log loss by 7% through feature engineering and multithreading techniques for Microsoft malware detection project.
- Improved accuracy achieved by adding ASM unigrams and ASM image features with bytegram unigrams
- Demonstrated expertise in feature engineering, multithreading, and machine learning algorithms, showcasing ability to handle large datasets and optimize performance.
- **https://github.com/bagh2/Microsoft-Malware-detection-problem-**

TRAINING / COURSES

Applied machine learning course

Codebasics from beginner to advance

python bootcamp from zero to hero