

# ANUJ BAGHEL

## DATA SCIENTIST

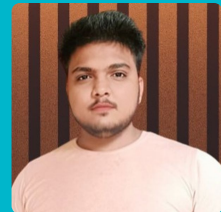
+91-9004672502

@ anujbagh300@gmail.com

<https://www.linkedin.com/in/anuj-baghel441518191/>

Mumbai, Maharastra

Github : <https://github.com/bagh2/Anuj-Baghel>



## PROJECTS

### Healthcare claim provider fraud detection

#### Machine-learning

01/2023 - 02/2023 Remote

<https://github.com/bagh2/Healthcare-claim-provider-frauddetection-Facebook>

- Healthcare Claim Provider Fraud Detection - ML REMOTEConducted a healthcare claim provider fraud detection project with Kaggle data,including 4 CSV files. Encountered highly imbalanced data with only 9% of providersbeing fraudulent. Used SMOTE and ADASYN to handle the imbalance problem. Triedvarious 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.
- Technologies used: Python, Sampling techniques: SMOTE, ADASYN, ML algorithms,ensemble techniques, neural networks, AWS.**

### Facebook Friend reccomandation

#### Reccomendation - System

11/2022 - 12/2022 Remote

<https://github.com/bagh2/Facebook-Friend-Recommendation>

- Successfully completed a Facebook friend recommendation project using Kaggle data provided by Facebook for research purposes, implementing various distance measuring techniques such as rank indexing, Jaccard distance, cosine similarity, and Adar indexing.
- Achieved high performance in both training and testing through the use of SVD factorization and preferential attachment techniques, training the model with different algorithms including Random Forest and XGBoost, and performing feature selection to identify the most important feature for friend recommendation, resulting in a test f1-score of 0.926 for Random Forest and 0.927 for XGBoost.
- Techniques utilized include rank indexing, Jaccard distance, cosine similarity, Adarindexing, Katz Centrality, Hits score, SVD factorization, preferential attachment,Random Forest, and XGBoost.**

### CUSTOM CHATBOT ON INDIAN CONSTITUTION

#### NLP

2023 - Present

- Developed a custom end-to-end chatbot on the Indian Constitution to make it easily accessible for simple users by downloading the official constitutional book from their website and performing end-to-end preprocessing, such as converting it into text format and removing unnecessary characters.
- Utilized various NLP techniques, including cosine similarity, named entity recognition, and graph-based search, to prepare the question and answer columns for training and trained the chatbot model using a pre-trained BERT model and tokenizer.
- Tested and evaluated the model using Dialogflow and deployed it to allow users to ask any question related to the Indian Constitution.
- Techniques used in this project: End-to-end preprocessing NLP techniques (cosine similarity, named entity recognition, graph-based search) Pre-trained BERT model and tokenizer**

## TECH STACK

PYTHON

MYSQL

DATA ANALYSIS

DATA VISUALIZATION

MACHINE LEARNING

NLP

RECCOMENDATION -SYSTEM

LARGE LANGUAGE MODELS

SCIKIT-LEARN

TENSORFLOW & KERAS

COMPUTER VISION

AWS & FLASK

## KEY ACCOMPLISHMENTS



**Implement and solve the Microsoft malware detection problem by reducing the log-loss by less than 1%.**

Handle the dataset larger than 200 GB in Google Colab and improve the log loss by less than 1%. Optimize the performance using advanced machine learning techniques.



**Building an end-to-end custom chatbot to help users understand the Indian Constitution.**

Build a custom chatbot that enables users to understand the Indian Constitution by asking any question related to it, such as which section pertains to a particular crime or what their rights are. The chatbot will provide comprehensive answers on the Constitution, making it easily accessible to the average user.

## EDUCATION

B.Sc: INFORMATION TECHNOLOGY

Mumbai,Maharatsra

2019 - 2022

CGPA

9.3 / 10.0