Shubham Saurabh

Sophomore :Undergraduate(graduation year:2027)

Major : CSE(Data Science Specialization)

Professor Colony,Main Road,Begusarai,Bihar,India|(+91)9264132652|shubham.saurabh.ds27@heritageit.edu.in|

Academic Qualifications:

year	degree/certificate	institute/school	score
2023 - 2027	Computer Science and Engineering (DSc specialization)	Heritage Institute of Technology,Kolkata	8.42/10 SGPA
2019 - 2021	C.B.S.E(XII)	Vikas Vidyalaya,Begusarai	73.2%
2017-2019	I.C.S.E(X)	St. Paul's School,Begusarai	90.2%

Scholastic Achievements:

In Olympiads	Secured State Rank-19 and National Top1% in Regional Mathematical Olympiad(RMO) ,conducted by NBHM ,Govt. Of India.
At HITK	Secured rank 12 in hackathon named Hack Heritage for selection of SIH- 2024

Work experience:

Systems Intern – Indian Oil Corporation Ltd (IOCL) |

Engineered a scalable sentiment analysis pipeline using RoBERTa for IOCL's internal Twitter Dashboard project to process and classify public feedback in real time

- Fine-tuned RoBERTa to detect nuanced sentiment and auto-classify tweets by service category (fuel, billing, safety, etc.)
- Built a resilient pipeline integrating Twitter API, preprocessing layers, and near real time data retrieval with sub second latency
- Enabled early detection of high-impact tweets and campaign sentiment trends, supporting rapid decision-making at scale across departments

Key Projects:

WATER FOOTPRINT CALCULATOR ML MODEL |

- Models: Linear Regression, Random Forest Regressor from sk-learn to check out best accuracy score and use best score algorithm for model
- Technical analysis:used KNN-Imputer to fill the null values in the dataset and use ridge regression from sk-learn to avoid overfitting in training data and used StandardScaler from sk-learn to have a unit variance and used PCA from sk-learn to check outliers and clusters
- Performance:achieved accuracy of around 95 - 98 percent on test data without underfit or overfit issue

BERT-BASE-UNCASED DISEASE PREDICTOR ML MODEL |

- Aim :To fine-tune a BERT-based symptom classification system using Hugging Face(pre-trained model) to predict diseases based on symptoms with a confidence score
- My work : Designed an efficient tokenization pipeline using AutoTokenizer to preprocess multi-class symptom datasets for transformer-based classification,using gemini API in the project to change paragraph symptoms to word oriented symptoms and have a good confidence score and if low score then some other disease can be present

Technical skills:

Algorithmic:python(expert),C++(intermediate),C(prior experience)Additional:MySQL,postgres,Git,kafka,docker(begineer)
ML:TensorFlow,PyTorch,Scikit-Learn,StableBaselines,Numpy,Pandas,Plotly,Matplotlib,MATLAB,SciPy,mechanistic interpretability(begineer)