

# RUTVIK KUMAR

## DATA SCIENTIST

### CONTACT



+91-9157638628



Ahmedabad, Gujarat, India

### SKILLS

Python                      Machine learning  
Deep learning              Computer Vision  
Artificial Intelligence              NLP  
Problem Solving                      DSA

### LIBRARIES

Sklearn,              NumPy,              Pandas  
Matplotlib,              SciPy,              Keras  
TensorFlow,              Streamlit,              PyTorch  
NLTK,              Hugging Face,              OpenCV

### DATABASE

MySQL ORACLE    BigQuery    EXCEL

### MACHINE LEARNING

Linear              Regression,              Logistic  
Regression, Decision Tree, Random  
Forest, Gradient Boost, XGBoost,  
AdaBoost, SVM,    KNN,    K-means

### DEEP LEARNING

ANN,              CNN,              RNN,              LSTM,  
Bi-directional    LSTM,              Encoder-  
Decoder,              Attention              Model,  
Transformer,              BERT,              VGG16,  
ResNet50, Inception\_V3, YOLO

### CERTIFICATIONS

GenAI with LLM (Deeplearning.AI)  
Advanced SQL (HackerRank)  
Python (HackerRank)  
Problem Solving (HackerRank)

## EDUCATION

### B. TECH (DUAL)

IIT BHUBANESWAR, MAY 2024  
CGPA- 8.01

### CBSE(XII)

APS, MAY 2019  
85%

### CBSE(X)

K.V, APRIL 2017  
CGPA-9.2

## EXPERIENCE

### HACKATHON

TREDENCE  
Beyond Possible

TREDENCE

MAY 2023

- Secured 2<sup>nd</sup> position, conducted by TREDENCE at Pan India level.
- Solved **Advanced NLP** problem of classification.
- Performed EDA on text data using **WordCloud** and **HeatMap**.
- Used **Bi-directional LSTM RNN** and LLM like **Distilbert-base** and then evaluated using **confusion matrix** and **accuracy V/S epoch curve**.

### DATA SCIENCE INTERN



GOJEK, BANGALORE

MAY 2022 – AUG 2022

- Worked with the GOFOOD team on the **Tensoba** project to predict food preparation time (FPT) for accurate prediction of ETA.
- Extract real-time and historical data using **BigQuery** and perform **EDA**, **data cleaning**, **Data Transformation** and **Feature Engineering**.
- Applying the **NLP** method like **TF-IDF** and **Word2vec** to convert dish names and using **PCA**, **T-SNE** and **RSS** for dimensionality reduction.
- Algorithms like **XGBoost** regression and classification (after Bucketing the FPT) are used to predict the FPT and reduce the **MAPE** by 5%.

### DATA SCIENCE/ANALYST INTERN



HENRY HARVIN ANALYTICS

JAN 2022 - FEB 2022

- Identifying flower images into different classes.
- I used **MobileNet\_v2**, **AlexNet**, **VGG16**, **VGG19** and a custom sequential CNN model using **Padding**, **Pooling** and **dropout layer** for prediction.
- Evaluate using **Accuracy Score** and **Precision**, deploy in **STREAMLIT**.

## PROJECTS

### BIGBERT WEB APP

- LLM-based web application on question answering where we upload the image and ask questions related to text in the image.
- Used genism models like **Word2Vec**, **Fasttext** and **GoogleNews-vector-negative300** for word embedding and pre-processing of text on **SQUAD** dataset for ML models like **Gradient Boosting** and **SVR**.
- Used the **BERT** model like **distilbert-base-cased-distilled-squad** and **deepset/roberta-base-squad2** and compared it with the ML model.
- Deployed the app on **streamlit.io** and made it public.

### CHEST CANCER DETECTION

- Classifying **CT-Scan** image into 3 types of cancer or NO-cancer.
- Performed **Data Augmentation**, **Rescaling**, and analyses of dataset **statistics** and removed **class imbalances** by assigning weights.
- Used **Inception\_V3**, **ResNet 50**, **VGG16** and **MobileNet** model for prediction and **Accuracy**, **F1-Score** and **Precision** for evaluation.