Kushal Kumar

DATA SCIENTIST

9708754348 | kushalarya00@gmail.com | Hyderabad, India

Objective

Experienced IT professional with a 4+ year track record, including 2+ years as a data scientist. Proficient in using data to solve complex problems and drive business decisions. Strong in programming, statistics, and data visualization. Effective communicator and collaborator, capable of working with cross-functional teams. Committed to continuous learning and ethical data practices. Eager to take on new challenges and contribute to organizational success.

Experience

Sysark Datasol pvt ltd | Data Scientist   01 / 2022 - Present

- Collaborate with data engineers and data wranglers to gather and transform data.
- Clean and preprocess data to ensure its quality and suitability for analysis.
- Explore data and perform feature engineering to create informative variables.
- Conduct data analysis to extract insights and trends from large datasets.
- Work on more advanced machine learning techniques, including deep learning, natural language processing, and reinforcement learning.
- Effectively communicate findings and insights to non-technical stakeholders.

Sysark Datasol pvt ltd | Associate Software Engineer   09 / 2019 – 12 / 2021

- Write, test, and maintain high-quality code for software applications or systems.
- Analyze and debug software issues and defects, providing effective solutions in a timely manner.
- Collaborate with senior engineers and architects to learn best practices and improve coding skills.
- Stay updated on the latest technologies and development practices to enhance your skills.
- Take part in training and workshops to further your professional growth.

Education

Data Trained Education Pvt. Ltd   01 / 2021 – 01 / 2022

PG Program in Data Science, Machine Learning and Neural Network | New Delhi

B. Tech (EEE)   10 / 2011 – 10 / 2016

R.P. Sharma IT | Patna

Skills & Tools

- Python
- Statistics
- Machine Learning
- Pandas

- NumPy
- Matplotlib
- Scikit-Learn
- Seaborn
- TensorFlow

- Keras
- PyTorch
- OpenCV
- MediaPipe
- Deep Learning

- Computer Vision
- Classification
- Regression
- Clustering
- Decision trees
- Random forests
- K – Nearest Neighbours (KNN)
- Support Vector Machine (SVM)
- XGBoost
- AdaBoost
- Gradient Boost
- RNN
- ANN
- CNN
- GAN
- NLP
- NLTK
- LSTM
- Bert
- R Programming
- PySpark
- Hadoop
- Azure ML
- Flask
- MySQL
- Tableau
- Excel

Projects

Bank Loan Default Risk Analysis

- **Objective:** The aim of this project is to help bank to make informed lending decision, mitigate risk and ensure the stability of their loan portfolios. In this project we develop robust model that can help optimize loan approval processes, improve risk management strategies and maintain financial sustainability.
- Solution Details:
- **Data Cleaning and Preprocessing:** Gather a comprehensive dataset of loan applications, historical loan data, credit reports, financial statements, and other relevant information. Cleanse and preprocess the data by handling missing values, outliers, and inconsistencies.
- **Feature Engineering and Selection:** Engineer informative features from the available data that can provide insights into loan default risks. Select relevant features that have a significant impact on loan defaults.
- **Data Analysis and Exploration:** Conduct exploratory data analysis to understand the distribution, correlations, and patterns within the dataset. Identify potential relationships between features and loan defaults, uncover trends, and gain insights into factors that contribute to default risks.
- **Model Selection and Training:** Choose appropriate machine learning algorithms or statistical models for loan default prediction. Train the selected models using the preprocessed data and evaluate their performance.
- **Model Evaluation and Performance Metrics:** Assess the performance of the trained models using suitable evaluation metrics such as accuracy, precision, recall, F1-score. Compare the performance of different models to identify the most effective one for loan default risk analysis.
- **Achievement:** Did Random Forest and achieved a 52% accuracy score.

Malignant Comment Classification

- **Objective:** The objective is to create an NLP model that identifies and flags malignant comments in online platforms, promoting a safer and more inclusive online environment by automatically detecting harmful content.
- **Solution Details:**
- **Data Collection and Annotation:** Gather a large dataset of comments or text data from online platforms that includes both malignant and non-malignant comments. Annotate the dataset by

labelling comments as either malignant (toxic, abusive, hate speech) or non-malignant (non-toxic, respectful).

- **Data Preprocessing:** Cleanse and preprocess the text data to remove noise, irrelevant information, or special characters. Perform techniques like tokenization, stop word removal, stemming or lemmatization, and handle issues such as misspellings or abbreviations.
- **Feature Extraction and Representation:** Convert the text data into numerical features that machine learning algorithms can process. Utilize techniques like bag-of-words.
- **Model Training and Evaluation:** Train an NLP model Naïve Baye using the labelled dataset. Split the data into training and testing sets. Evaluate the model's performance using appropriate metrics like accuracy, precision, recall, F1-score.
- **Achievement:** Did Naïve Baye and achieved an 61% F1 score.

Pose Landmark Detection

- **Objective:** Develop a real-time computer vision model to accurately detect and track body landmarks for applications like fitness tracking, augmented reality, and gesture recognition.
- Solution Details:
- **Data Collection:** Gather a diverse dataset of images or video footage containing human subjects in various poses and movements. Include a wide range of body types, clothing styles, and backgrounds to ensure the model's robustness.
- **MediaPipe Integration:** Utilize the MediaPipe library, which provides pre-trained models and an efficient framework for real-time pose estimation. Integrate MediaPipe into your project to leverage its powerful pose landmark detection capabilities.
- **Model Fine-tuning:** Fine-tune the pre-trained pose landmark detection model from MediaPipe using your collected dataset. This step helps to adapt the model to the specific context of your application and improve its accuracy.
- **Real-time Processing:** Implement the necessary algorithms and techniques to enable real-time pose landmark detection using MediaPipe. This involves efficiently processing frames or video streams, detecting and tracking the pose landmarks in each frame, and visualizing the results in real-time.
- **Evaluation and Optimization:** Optimize pose landmark detection model by experimenting with techniques like data augmentation, model architectures, and hyperparameter tuning to improve performance, using metrics like mean average precision (mAP) and accuracy for evaluation.
- **Achievement:** Implemented real-time Pose Landmark Detection using MediaPipe with high accuracy. Enables fitness tracking, augmented reality, motion analysis, and gesture recognition. Rigorous evaluation and optimization ensure reliability.

Face and Eye Detection from webcam

- **Objective:** The objective of this project is to develop a computer vision system that can detect faces and eyes in real-time using a webcam. The system should accurately identify the locations of faces and eyes within the video stream and provide visual feedback to the user.
- **Solution Details:**
- **Capture Webcam Feed:** Access the webcam and continuously capture video frames.
- **Preprocess Frames**: Apply necessary preprocessing steps such as resizing, converting to grayscale, or normalizing pixel values to enhance the quality of the frames.
- **Face Detection:** Utilize a pre-trained face detection model, such as Haar cascades, to identify the regions of interest where faces are likely to be present in each frame.
- **Eye Detection:** Once a face is detected, apply a pre-trained eye detection model, such as Haar cascades or deep learning-based models, to identify the regions of interest corresponding to eyes within the detected face region.

- **Visual Feedback**: Draw bounding boxes or overlays around the detected faces and eyes on the original frame to provide visual feedback to the user. This can be done using OpenCV's drawing functions.
- **Real-Time Display:** Continuously display the processed frames with the detected faces and eyes in a separate window or on the screen.
- **Achievement:** Completion of this project delivers real-time face and eye detection for video surveillance, driver drowsiness detection, and user-computer interaction. The system provides accurate real-time feedback, highlighting the potential of computer vision and OpenCV in object detection.

Certificates

- Microsoft Certificate AI.
- Overview of Geoprocessing using Python Certificate by ISRO.
- Python Certificate by Skillzcafe.
- Machine Learning Certificate by IBM.
- Excel Chart and Graphs Master Data Visualization Certificate by Skillzcafe.
- IoT 4.0 Certificate by TATA Steel.
- Fundamentals Data Analysis & Decision-Making Models Theory Certificate by Skillzcafe.
- Microsoft Excel - Basic Excel Formulas and Functions Certificate by Skillzcafe.
- Advance Excel Certificate by TATA Steel.

Interests/ Hobbies

Learning new things, Cooking, Watching Movies etc.

Personal Details

Date of Birth : 28/02/1993

Languages : Hindi, English

Address : New Mainpura, Kharanja Road, Near Sapt Devi Mandir, Saguna More, Danapur Cantt, Patna, Bihar – 801503.