https://www.linkedin.com/in/harshal-patil-annu

HARSHAL PATIL

9730942181

Harshalpaatil20@gmail.com

Data Scientists

AT Post Kapadane dhule 424307 MAH

https://github.com/HarshalPatil20

SUMMARY

As a Data Scientist, I am a skilled professional with expertise in Python, MySQL, Machine Learning, Excel, Data Analysis, Data Cleaning, Data Mining, EDA, and Statistics. With a strong academic background in Electronics and Telecommunication Engineering, I have successfully completed projects such as analyzing employee performance for HR analytics, analyzing Spotify and YouTube songs, and analyzing user behavior on Instagram. Additionally, I have completed a business analytics internship where I gained practical experience using Python, Pandas, NumPy, Matplotlib, Excel, and SQL. I am passionate about leveraging data-driven insights to drive business growth and improve decision-making processes. With a commitment to continuous learning and a detail-oriented approach, I am ready to contribute to real-time data science challenges.

EDUCATION

Sandip Foundation SITRC Nashik

Bachelor's Degree in E&TC
2017 – 2021 CGPA - 7.61

S.S.V.P.Science College Dhule

BHSC 2016– 2017 Percentage : 72.15%

SKILLS

- Python , SQL , Statistics
- Machine Learning, Deep Learning
- Regression (Logistic, Linear)
- Random Forest, Bagging
- XG-Boost, GBM, Light GBM, Cat Boost, Boosting
- NLP, NLG, BERT
- Image Processing, Computer Vision
- Deep Learning, CNN,RNN
- Resnet, Transformers
- Tensor Flow, Keras
- Pytorch.
- Exceptional communication and interpersonal skills
- MLOPS , GIT, Creating CICD pipeline
- Dockers, Kubernetes
- Excel, Power BI, Power Point

CERTIFICATIONS

- Python ,
- Data Science
- Machine Learning
- Data analyst

EXPERIENCE

Project Engineer

Wipro Company | 2021 - Present

- Developing new Portals, Pages, and Workflow, and Creating an Automatic
- Testing Framework for ServiceNow Application.
- Automated the Change Management, ITSM, CMDB, Service Portal etc

PROJECT

Tomato plant leaves disease detection using Image processing.

Institute Project 2021

- Plant leaf disease is a significant problem in today's agriculture industry. This project aims to detect and identify leaf diseases and provide preventive measures to farmers using Image processing methodology.
- The objective of this project is to identify plant leaf disease and classify it into different types, suggesting appropriate pesticides for each disease.
- The system dataset contains 14,000 images of tomato leaves with different types of diseases such as bacterial blight, mosaic virus, leaf mold, etc. The approach involves several steps, including image acquisition, image preprocessing, image segmentation, feature extraction, and image classification/model building.
- For model building, we use the CNN algorithm equipped with input, output, and hidden layers, all of which help process and classify images. The hidden layers consist of convolutional layers, ReLU layers, pooling layers, and fully connected layers, which play a crucial role.
- The infected region of the leaf is segmented and analyzed, and the image is fed to our application for identification of diseases.
- This project provides a valuable tool for the agriculture community, especially in remote villages, and minimizes human errors.
- Skills:- Python, ML, Imagine Processing, Deep Learning, CNN

Stores Sales Prediction (2023)

- Optimizing inventory management and accurately forecasting client demand are crucial for shopping malls and Big Marts. The vast amount of consumer information and item details stored in data warehouses presents an opportunity to uncover valuable insights. By leveraging data mining techniques, we can identify anomalies and discover common patterns that can enhance inventory management strategies and improve sales forecasting
- Using a comprehensive machine learning approach, I performed data exploration, cleaning, feature engineering, model building, and testing. Leveraging the XGBoost algorithm, I developed a tailored solution for predicting sales in different Big Mart stores, achieving an 85% accuracy rate.
- Data is collected from various resources, and as part of the analysis process, I identified missing values, cleaned the data, and utilized data visualization techniques to enhance understanding.
- For deployment, I utilized the Flask framework to create an application, and for easy accessibility, I developed a simple webpage. The deployed model is hosted on a cloud platform, allowing seamless access for everyone involved in handling the application.
- Skills: Python, Machine Learning, Flask, HTML, SQL

Food Delivery App Data Analysis

- This project focuses on analyzing a real-world database using SQL, extracting valuable insights, preprocessing the data in Python for improved performance, and utilizing structured query language to retrieve useful information.
- The main objectives include cleaning and analyzing the dataset, mining insightful information, and visualizing the data to gain a better understanding of restaurant performance. Key Business Problems Solved:
- 1. Determining the most popular dish from the highest-rated restaurant on Zomato, which offers online ordering and delivery services.
- 2. Identifying the top 15 restaurants with a minimum of 150 votes, a rating greater than 3, and no current online ordering, to assist Zomato in expanding its business and maximizing profit. The restaurants with the highest votes are prioritized. By addressing these business problems, this project contributes to enhancing decision-making and strategic planning within the restaurant industry.

Analyzing Employee Performance for HR Analytics Using Python and MySQL

- Data was collected from HR resources and underwent a data cleaning process. Duplicate rows were removed, and rows with irrelevant data types for numeric columns were also removed. Any irrelevant values in each column were also removed, and all values in each column were validated to check for any inconsistencies or discrepancies in data types, units, or formats.
- Insights were derived from the HR data, such as the number of trainings, previous year rating, length of service, and average training score. These features were found to affect employee behavior towards work.
- The average training score of employees was 63.17, and the average rating of employees was 3.3. Most employees (301.95%) were working in Sales & Marketing, while 20.63% were in the Operations department. The remaining percentage of employees worked in different departments: 4.86% in HR, 4.70% in Finance, 1.94% in Legal, and 1.94% in R&D.

ACHIEVEMENT

Best Paper for the paper entitled Tomato Plant leaf disease detection using image processing in 7th National Conference Paper Publication:- IOSR Journal of Engineering (IOSRJEN) www.iosrjen.org ISSN (e): 2250-3021, ISSN (p): 2278-8719 Special Issue || September 2021 || PP 26-28 7th National Conference On Advancements in Communication, Computing and Electronics Technology-ACCET 2021

LANGUAGE

ENGLISH, MARATHI, HINDI