Karan Malhotra

Data Engineer | Analytics & Cloud Solutions | Machine Learning | Python, SQL

Gurugram, India 8095631278 malhotrak1999@gmail.com https://web-portfolio-rosy.vercel.app

SKILLS

Programming
Python, C, C++

SQL, Shell, HTML, Javascript,

Java, CSS, C#

Frameworks and Tools
Jupyter Notebook, PostgreSQL, VS Code, Google Colab, Kaggle Kernel, Heroku, AWS, GCP, Azure Blob Storage, Azure Data Factory, Azure Databricks, Azure Kubernetes, Data Studio, Git, Firebase, Linux, Power BI, Pycharm, Nodejs, Reactjs

ML Framework
Scikit-learn, Keras, Pandas, Numpy, Matplotlib, Seaborn, Flask, NLTK, Tensorflow, Streamlit, PyTorch, XGBoost, LightGBM, Hugging Face Transformers, Dask, Spark, Apache Airflow

Miscellaneous
App Development, Data Warehousing, Data Analysis, Data Modelling, Automation, Project Management, Agile Methdology

CERTIFICATIONS

Overall SQL for DS Challenge
AlmaBetter

Full Stack Data Science
AlmaBetter

Advanced Machine Learning
AlmaBetter

Master AI
Udemy

Advanced Analytics Framework
AlmaBetter

MS Excel A-Z
Udemy

Software Development Processes and Methodologies
Coursera

ACHIEVEMENTS

Gold Badge in Python & SQL

HackerRank || 550 Points in Python || 820 points in SQL

HackerRank

INTERESTS

Cricket

Reading Books

Gaming

PROFILES

kmalhotra657
LinkedIn

Vagueken
Github

malhotrak1999
Medium

WORK EXPERIENCE

IraCommerce
Data Engineer
Hyderabad(Remote)
June 2023 - May 2024

- Developed and **optimized** web scraping scripts using the **Scrapy** framework to extract structured product data from diverse e-commerce platforms such as Amazon, Flipkart, Walmart. Boosted **scraping speed** by 40% and enhanced data retrieval efficiency.
- Implemented Azure Kubernetes Spark Integration **(AKS on Spark)** to run **ML models**, reducing dependency on Azure Databricks (ADB) by **50%**. Automated data reporting, saving **18 hours**, and perfected data visualization for daily data.
- Leveraged **Azure Data Lake** to store and manage large volumes of structured and unstructured data, achieving **99.9%** data availability and integrity with regular backups.

Website

AlmaBetter
Data Analyst Trainee
Bengaluru(Remote)
October 2022 - May 2023

- Acquired competencies in **Feature Engineering, EDA, Power BI, Tableau, ML.**
- Mastered **Python** and **SQL** through diverse projects, demonstrating a strong aptitude for Data Manipulations.
- Secured a place in the **top 5%** of students in the cohort of **400** students, and worked as an **SME.**

Website

EDUCATION

IBS Hyderabad
B.Tech Computer Science
7.1 CGPA
Bachelors
March 2017 - April 2021

Coursework: Deep Learning, Internet Applications, Security and Privacy, Digital Image Processing, Microprocessors, Software Design Analysis, Information Management, Digital System Design, Management for Engineers, Data Structures, Optimization Techniques, Engineering Design, Operating System.

Website

PROJECTS

Online Workplace Visual Sentiment Detection

- Attained **88% accuracy** with **Keras CNN** and **Transfer Learning MobileNetV2** models, reducing training time and improving accuracy.
- Boosted model accuracy by **16%** through **image augmentation** and **early stopping** techniques.
- Deployed a quantized model through **Streamlit API** for conducting sentiment analysis, resulting in a **20%** boost in overall satisfaction scores. **Link to the dashboard -** "Facial Emotion Recognition".

Image Preprocessing, Deep Learning, Neural Networks, Computer Vision, Transfer Learning, Model Deployment

Github

Netflix Movies and TV Shows Clustering Prediction

- Achieved optimal clustering with **0.56 average silhouette score for 5 clusters** using **elbow method** and **silhouette score analysis**.
- Improved NLP model performance by 25% through essential text preprocessing techniques, such as **stopword removal, stemming, PCA**, and NER.
- Augumented ML model **accuracy by 30%** through text data clustering and built a Streamlit recommendation system that boosted user engagement by 30%. **Link to the dashboard -** "Netflix Recommender System App".

Multi-Class Classification, NLP, NER, Unsupervised Learning, TF-IDF

Github

Company Bankruptcy Prediction

- Experimented with **six ML models** to achieve a **95% classification accuracy**, resulting in improved accuracy and reduced training time.
- Increased model adoption by **20%** by leveraging Shap plot values to **identify feature importance** and **make data-driven decisions**.
- Fine-tune **deployment** efficiency by saving the **best-performing ML model**, **XG Boost**, in a pickle file for future use.

ROCAUC, F1, Shap, Random Forest, XG Boost, Pickle Dump

Github