Shubham Vishnu Khedekar

+91-9923690696 | shubhamkhedekar2206@gmail.com | GitHub | LinkedIn

SUMMARY

Enthusiastic Data Scientist with 9 months of experience in predictive modeling, algorithm optimization, and data-driven solutions. Proficient in Python (TensorFlow, PyTorch), SQL (MySQL), Tableau, and Power BI, with expertise in image processing, data analysis, data manipulation, and visualization, and skilled at communicating complex insights effectively.

SKILLS

- Exploratory Data Analysis
- Data Analysis
- Data Visualization (Python Libraries, Power BI, Excel, Google Sheets)
- Deep Learning (ANN, CNN, RNN)
- Natural Language Processing (LSTM, GRU)
- Image Classification (CNN, YOLO, Holistic Model)
- LLM (Langchain, Langgraph)
- Model Deployment
- Text Classification (TIDF)
- Data Engineering
- Neural Networks
- Predictive Analysis
- SQL Query Development and Optimization
- Database Backup, Recovery, and Security
- Data Wrangling
- Regression Analysis
- Machine Learning (Regression, Classification, Clustering)
- Statistics and Mathematics
- Image Preprocessing (OpenCV)
- Time Series Analysis (LSTM)
- Generative AI
- Transformer (Hugging Face)
- RASA Chatbot
- Algorithms
- Pattern Recognition
- Version Control (Git)
- ETL Processes and Data Integration
- Database Programming and SQL

Programming Language: Python, SQL

Software: Python, Power BI, MySQL, VS Code, PyCharm, Git and GitHub, Excel, Power Point

Libraries: NumPy, Pandas, Matplotlib, Seaborn, Streamlit, Scikit-Learn, Tensorflow

Work EXPERIENCE

Machine Learning Intern | Cognition Technology August 2024 – February 2024

- Developed an end-to-end machine learning project in the insurance domain, including data preprocessing, model development, evaluation, and deployment.
- Built a CI/CD pipeline using GitHub Actions for automated model build and deployment; hosted the solution on Heroku cloud environment. Implemented and experimented with multiple ML and Deep Learning algorithms to solve diverse project use cases.
- Collaborated with cross-functional teams to align data strategies with business objectives and documented workflows for reproducibility.
- Proactively enhanced skills by researching and applying the latest data science methodologies and tools.

Data Analyst Intern | Acmegrade Industry February 2024 – April 2024

- Compiled data from client in raw excel format or from sharepoint; increased analysis by 50% by using pivot table & Vlookup.
- Designed powerbi dashboard for financial claims portfolio of 200+ client accounts to review the KPI; results in 25% improvement in collection rates and 30% reduction in claim denial rates, leading to faster revenue recognition
- Directed teams in coordinating efforts to address billing issues, facilitate payment discussions, and achieved a 15 % improvement in client satisfaction.
- Utilized SQL for data extraction, aggregation, and querying to generate reports for business intelligence.

- Collaborated with senior teams to develop analytical solutions and prepared comprehensive summary reports for projects.

EDUCATION

Master of Computer Application, Pune India August 2021 – October 2023
MES's Institute of Management & Career Courses (IMCC)
Bachelor of Computer Science , Chh. Sambhaji Nagar September 2018 – June 2021
Deogiri Institute of Technology & Management Studies (DITMS)

Projects

Human Facial Emotions Detection | (Python, TensorFlow/Keras, OpenCV, NumPy, Matplotlib, CNN)
• Built a deep learning model using Convolutional Neural Networks (CNN) to classify facial expressions into multiple emotions (happy, sad, angry, neutral, etc.).
• Preprocessed image data (grayscale conversion, resizing, normalization, and data augmentation) for robust training.
• Designed and trained a CNN model with multiple convolution, pooling, and fully connected layers for feature extraction and classification. Achieved 70-75% accuracy (replace with your actual result) on the FER2013 dataset using TensorFlow/Keras.
• Developed a real-time emotion detection pipeline using OpenCV for live webcam video input.

Predictive Maintenance Model | (Data Preprocessing, Scikit-learn, Model Evaluation, Random Forest, Seaborn)
- Built a machine learning model to predict machine failures using sensor data (temperature, torque, rotational speed, tool wear).
- Preprocessed raw data (outlier removal, scaling, encoding) and handled class imbalance with SMOTE.
- Trained and evaluated multiple models (Random Forest, Decision Tree, Logistic Regression, SVM, KNN), achieving 90-95% accuracy / ROC-AUC score on unseen test data.
- Implemented failure-risk prediction for new sensor inputs to support maintenance scheduling and downtime reduction.

IPL Score Prediction | (Regression Model, Feature Engneering, Optuna, Streamlit, Data Analysis)
- Built a machine learning regression model to predict IPL final scores, achieving 95% prediction accuracy.
- Applied feature engineering and Optuna hyperparameter tuning, boosting model performance by 20%.
- Developed an interactive Streamlit web app allowing users to predict scores from live match data, improving user engagement by 35%.
- Analyzed 1M+ historical IPL data points, identifying key features influencing match outcomes for data-driven team strategies.

CERTIFICATION

- Master in Data Science & Analytics with AI
- IBM Data Analytics Professional
- Programming in Python
- Power-BI Data Analyst
- Hacker Rank SQL