ASHUTOSH SUSARE

Data Scientist

CONTACT

📞 +91 8788266214

✉ ashutosh.susare@gmail.com

📍 B-003, Prakash Vivacity , Thite Nagar, Kharadi, Pune, Maharashtra 411014

EDUCATION

Bachelor of Engineering (BE)
Pune University
2020 | 9.23 CGPA

Diploma
MSBTE
2017 | 91.00 %

HSC
Pune Board
2015 | 60.00 %

MACHINE LEARNING

- Linear Regression
- Logistic Regression
- Ridge & Lasso Regression
- Naïve Bayes Classifier
- k Nearest Neighbor's Classifier (KNN)
- Support Vector Machine (SVM)
- Decision Tree
- Random Forest
- Gradient Descent
- Ada-Boost
- Gradient Boosting, XG-Boost
- K-means Clustering.

PYTHON / ML PACKAGES

- Pandas
- NumPy
- SciPy
- Scikit-learn
- Seaborn
- Matplotlib
- Flask APIs

PROFILE

Experienced Data Scientist with 4+ years leveraging data analysis, statistics, and machine learning for actionable insights. Proficient in Python, skilled in preprocessing, predictive modeling, and visualization. Passionate about solving complex problems and driving data-driven decisions through ML, DL, & NLP. Strong communicator and collaborator.

WORK EXPERIENCE

- A professional with experience in Artificial Intelligence, Data Science, Machine learning, and NLP with expertise in Financial and Healthcare domain projects.
- Able to investigate Data Visualization and summarization techniques conveying key findings. Object Oriented Programming (OOP) in Python.
- Experience in Web Framework Flask and Cloud Platform. Thorough understanding of Probability and Statistics, Bayesian methods, Time Series analysis, Non-Relational databases.
- Ability to Designing and programming ML methods and integrating them into our ML framework / pipeline.
- Source code management and Version Control system using Git and GitHub.
- Capability to handle big data and process image data through computer vision.

- **Data Scientist**
  MAS Financial Services Limited   *08/2023 - Present*

  Project 1 : Data Extraction and Aadhar Card Masking

  - Enhanced image readability & fixed orientation using PIL & OpenCV.
  - Developed a function to accept .png, .jpeg, .jpg, .tif, .tiff, and .pdf formats. Cropped images with multiple documents and identified document types using OCR.
  - Split PDFs with multiple documents by page using pdf2image.
  - Applied OCR for data extraction, utilizing regex and validation methods to improve accuracy.
  - Created a Flask API for integration with MAS LOS and deployed the project on AWS.

  Aadhar Masking

  - Connected to data sources and gathered Aadhar cards in various formats. Used RRDN for image super-resolution to enhance feature extraction.
  - Implemented a checksum function to validate Aadhar numbers.
  - Used regex to find Aadhar numbers and plotted bounding boxes with OpenCV. Applied Gaussian blur with OpenCV to mask Aadhar numbers.
  - Handled image rotation and features using PyTesseract. Deployed the server and connected to the database.

TECHNICAL SKILLS

Deep Learning :
Neural Network, Deep learning, ANN, CNN, RNN, LSTM, TensorFlow 2.x, Keras, PyTorch, Activation & loss functions, optimizers, Padding, Pooling, Generative AI (GenAI)

NLP Libraries :
Text understanding, Text Pre-processing, Test classification techniques, Text clustering skills.

Libraries : nltk, gensim, spacy, googletrans.

Techstat : BOW, TFIDF, word2vec, doc2vec, keypharse extraction, PowerBI, Tableau.

Databases :
MySQL, MongoDB, RDS, CRUD operations, Subqueries, Window functions, Joins.

Operating Systems OS :
UNIX, LINUX, Windows, Ubuntu

AWS :
- Elastic Compute Cloud (Ec2)
- Sagemaker
- Notebook instance
- AWS container
- Simple Storage Services S3
- Lambda (serverless compute cloud)
- Deployment, Multi cloud environment

Time-Series Analysis :
AR, MA, ARIMA, SARIMA models for time series data analysis.

Cloud Platforms :
AWS, GCP, Celoins, Azure.

MATH & STATS

- Gradient Descent
- Probability
- Linear algebra
- Bayes Theorem
- Hypothesis Testing
- Filter, Wrapper
- Embedded Method
- P-Value, Z-Test, T-Test,
- Info -Gain Test

CERTICIATIONS

- Machine Learning Certification
- SQL certified by HackerRank
- Agile certification.

Project 2: RASA Chatbot for Customer Service
- Designed a RASA chatbot for various user services.
- Implemented forms and slot mappings for entity extraction.
- Set up intents, rules, and stories for response management.
- Created action files to validate forms and manage responses.
- Developed scripts to send acknowledgment emails to users.
- Stored leads and chatbot responses in the database.
- Integrated the chatbot with the website.
- Outcome: Successfully automated customer service interactions and managed user data efficiently.

CIBIL Scrub Data Analysis (2022-2023)
The CIBIL scrub data analysis for 2022-2023 provided insights that improved our market cap and guided our strategy to expand the business. Through demographic and behavior analysis, we successfully increased trades by 31% compared to the previous financial year.

- ### Data Scientist

Tata Consultancy Services (TCS)   *Domain - BFSI*   *03/2021 - 08/2023*

Project : Term Deposit Prediction Model

Description : Predict whether the client will subscribe to a term deposit, based on the analysis of the campaigns data performed by DB bank.

Roles and Responsibilities :
- Acquisition of usable data from valuable data sources.
- Exploratory data analysis to analyze and investigate data.
- Processing, cleansing, and validating the integrity of data to be used for analysis.
- Enhancing data by using data visualization & summarization.
- Build predicative models using various ML tools to predict the possibility.
- Evaluation and analysis of various ML and deep learning model on the basis of statistic to achieve the best predication.
- Deployment of end to end model in production server.
- Demonstrate and represent the outcomes in clear manner to Business owners.

Project : Sentiment analysis model (NLP)

Description : Developed a sentiment analysis model using NLP techniques to classify text data into positive or negative sentiment.

Roles and Responsibilities :
- Preprocessed textual data by removing the stop words, punctions, and performing tokenization and stemming.
- Applied feature extraction method like TFIDF and count vectorizer to convert text into numeric vector.
- Enhancing data by using data visualization & summarization. Implementation of various classification ML algorithms to predict possibility.
- To achieve long range dependency by model we Trained model by using RNN and use LSTM.
- Implemented Hyperparameter tuning, cross validation and early stopping, dropout as well to optimize the accuracy and better performance
- Achieved an accuracy 86% By using the RNN and LSTM.