Vanashri Dipak Mane

Data Scientist at LearningBrite, Bangalore

Professional Summary:

Open-minded Data Scientist with 3+ years of experience. Experience in executive data-driven solutions to increase efficiency, accuracy, and utility of data processing. Knowledge of gathering, cleaning, and organizing data. Proficient in predictive modelling, data pre-processing, Machine Learning Algorithms, Deep Learning Neural Networks as well as scripting language Python to solve Industrial problems. Advanced understanding of Algebraic, Statistical, and other analytical techniques. Highly organized, motivated, process-oriented and diligent with significant background in mathematics.

Address: 204, Acme Valencia, Near Devanshi Bazar,Bhumkarnagar bridge, Narhe, Pune.
Pin code :411041

Current location: Pune.

Contact: +91 9860014975

Mail: Vanashri14mane@gmail.com

GitHub: Vanashri14

LinkedIn: Vanashri Mane

Technical Skills:

Languages:	**Python**, C, C++, Java
Database:	**SQL**, Mongo DB,DB2, MySQL, Oracle, MS Access
Cloud:	AWS
Data Science Library:	**Numpy, Pandas, Matplotlib, Seaborn, Power BI, Tableau**
Machine Learning Algorithms:	**Linear Regression, Logistic Regression, Decision Tree, Naïve Bayes, Random Forest, SVM, ARIMA, SARIMA, Time series Analysis**
NLP:	**TF-IDF, Word2Vec, Glove, BERT, Transformer, Hugging Face, Spacy, NLTK, BERT**
GenAI LLM:	**Transformers, BERT, OpenAI API, Langchain, Llama models, Fine tuning**
Deep Learning Algorithms:	**ANN, CNN, RNN, LSTM**, GRU, **Keras**, **Tensorflow**, **Pytorch**, SHAP, **OpenCV**, **Yolo** v5
VCS:	Git, GitHub
Agile:	Jenkins, Jeera
Tools:	**PyCharm**, **VSCode, Notebook**

Professional Skills:

- Experience in building GenAI applications.
- Knowledge of AWS services like Lambda, Bedrock, S3, Connect, CloudWatch.
- Worked directly with clients to design and develop GenAI solutions, ensuring alignment with business needs
- Expertise in Data Analysis, statistics, programming and ML techniques.
- Expertise in Python programming language.
- Experience in creating mature Data science pipelines encompassing Data standardization, Feature extraction, model validation and optimization.
- Cleaned, merged and manipulated datasets and conducted feature engineering using Pandas
- Hands-on experience of Data Science libraries in Python such as Pandas, NumPy, scikit-learn, Matplotlib, seaborn, Keras, OpenCV, TensorFlow.
- Novice basic knowledge in Python framework (like Django, Flask etc)
- Experienced in Visualization Tools Like Tableau, PowerBI.
- Having experience in aws cloud and working on lambda functions is preferable.
- Experienced with Git and version control best practices
- Experience with scripting in SQL - extracting large data sets and design of ETL flows
- Experienced in Agile development
- Having strong knowledge of Data Structure and Algorithms
- Good problem-solving skills, Positive attitude, self-motivated and Quick learner

Employment History:

Data scientist at **LearningBrite, Bangalore** (:: Learningbrite Technologies Private Limited ::)

Period: Dec 2021 to Present.

Projects:

Project 1: AI-Powered Video Analytics Platform **June 2024 – May 2025**

Domain: Artificial Intelligence

Primary Goals: To develop an AI-powered video analytics platform capable of analyzing and processing video content.

- To automate the generation of metadata from video content. To provide automated transcription and translation services.
- To perform sentiment analysis on video content.
- To enable face detection, scene detection, and object detection within video content.
- To store all extracted insights in a MongoDB database. Also generate AI-powered voice summaries of the processed text.

Skills used: Data Pre-processing, Data extraction, Sentiment analysis, Object detection, Face detection, Meta data handling.

Roles and responsibilities:

- Collect and manage video datasets from various sources (e.g., surveillance memy, online platforms, custom datasets).

- Perform video pre-processing, including frame extraction, resizing, and format conversion.
- Handle noise reduction, colour correction, and other video quality enhancement techniques.
- Annotate video data for training supervised models (e.g., bounding boxes, scene labels).
- Worked with Mongodb,vector database-chromadb,emmbeddings,AWS cloud,Langchain,LLM.Having good knowledge of them.

Project 2: Advanced Predictive Modelling for Bank Deposit Subscriptions.
Client: Royal Bank of Canada (Canada) **Mar, 2023 – Feb,2024**
Domain: Finance
Primary Goal: To predict whether a customer avails term deposit or not, also average portrait of the person who is willing to subscribe to a term deposit.
Skills used: Data Pre-processing, Data Visualization, EDA, Model training and evaluation using a supervised machine learning algorithm
Roles and responsibilities:

- Creating a proper record of each customer and clustering all of them to make it useful for further use with help of API
- Analysing the customer's data as per the records provided
- Dive deep into data, do analysis, and discover patterns and root causes. Generate insights that drive the product.
- Choosing important features for the further analysis as per client's specifications and 5C principal
- Building binary classifiers based on machine and deep learning models on real data in predicting loan default probability
- Take feedback from analysis, end-users, and domain experts to perform model calibration, bug fixes, and enhancements.
- Work on new client on boarding by developing configuration for model pipelines.
- Work on model delivery to the Production deployment team and coordinate model production deployments.
- Keep abreast of machine learning and industry developments, conducting R&D to incorporate best-in-class modelling methodologies and disseminate learning within the team.

Project 3: Clinical Text Classification using Convolutional Neural Networks
Client: United Health Group (USA) **Dec, 2021 - Nov, 2022**
Domain: Healthcare (Medical data)
Primary Goal: To develop an advanced clinical text classification system that combines the power of rule-based features and knowledge-guided convolutional neural networks (CNNs).
Skills used: – Text Preprocessing, Word Embeddings,Data Annotation,Data Cleaning,Model Building,Transfer Learning,Evaluation Metrics, Medical Domain Knowledge,Data Privacy and Compliance Data Exploration, Feature Engineering, Data Visualization, EDA, Model Building, pre and post-performance checking
Roles and responsibilities:

- Data Preprocessing: Clean and preprocess clinical text data, which involves tasks like lowercasing, removing special characters, and tokenization, implement text normalization and stemming if needed.
- Data Labeling: Annotate clinical text data with the relevant categories or labels, Ensure data consistency during the annotation process.
- Feature Engineering: Extract and transform text features suitable for CNN input, such as word embeddings or numerical representations.
- Model Development: Assist in building and training CNN models for text classification, contribute to the architecture design and hyper parameter tuning.
- Model Evaluation: Participate in evaluating model performance using appropriate metrics (e.g., accuracy, precision, recall), collaborate with senior data scientists to refine the model based on evaluation results.
- Data Visualization: Create visualizations to explore and communicate insights from the clinical text data.
- Documentation: Document data preprocessing steps, model development, and findings for team reference, ensure documentation is clear and accessible to other team members.
- Collaboration: Work closely with cross-functional team members, including data engineers and domain experts, participate in team meetings and share insights and progress.
- Quality Assurance: Conduct quality assurance checks to maintain data quality and integrity, collaborate with senior data scientists to improve quality assurance processes.
- Problem Solving: Actively engage in problem-solving, both independently and as part of a team,seek guidance and mentorship from senior team members when needed.

Education:

ME. Computer Engineering (SPPU) -8.6CGPA Completed in 2015.

Online Courses & Certifications:

- Complete Machine Learning Course with Python-Udemy
- Learn the building blocks of python for absolute beginners-Udemy
- Tableau Training-Skillup by Simplilearn
- Deep Learning for Beginners- Skillup by Simplilearn
- Data Science Tools- Cognitive Class
- Python for Data Science- Cognitive Class
- Introduction to Data Analytics- Skillup by Simplilearn
- Power BI for Beginners- Skillup by Simplilearn
- Python Libraries for Data Science- Skillup by Simplilearn