ABHIJEET GODE

• 8407906707 • Pune, Maharashtra • abhijeetgode1@gmail.com • LinkedIn • GitHub

Senior Software Engineer (Ai)

Machine Learning | Deep Learning | Natural Language Processing

Professional 4.9 years of dedicated experience within the IT industry. My expertise spans across Data Science, Machine Learning, Deep Learning, NLP, and Cloud technologies working on LLM, Generative AI. Proficient in Python, SQL, Linux and adept at crafting end-to-end ML solutions, I specialize in data preprocessing, feature engineering, model selection, tuning, and evaluation. Notably, I have leveraged the Flask framework for developing web-based applications. My track record reflects a commitment to designing and developing impactful ML models that directly addressorganizational objectives. As a collaborative team member, I bring strong problem-solving abilities and effective communication skills to the table.My passion for remaining abreast of emerging tech trends underscores my dedication to contributing to innovative data-driven solutions.

WORK EXPERIECE

❖ **Senior Software Engineer (AI Engineer) – Expleo Solutions** - *Pune, Maharashtra*   Aug 2024 – Sep 2024

As a Senior AI Engineer, I provided strategic AI-driven solutions, engaging in client use cases and consulting on innovative AI implementations, including:

- **Low-Code/No-Code Data Accelerator:** Developed a platform that enables businesses to rapidly process, analyse, and model data, providing insights that align with their operational needs.
- **Bank Fraud Detection System:** Built a rule-based database and machine learning framework to predict unusual transaction behaviours, improving fraud detection efficiency.
- **KYC AI Solution:** Created an AI-powered chatbot solution for seamless customer onboarding, utilizing machine learning to enhance Know Your Customer (KYC) compliance.
- Leveraged cutting-edge technologies such as Prompt Engineering, LangChain, Llama, OpenAI GPT-3.5 Turbo, Hugging Face pretrained models, and vector databases to deliver AI solutions.

❖ **Senior Data Scientist - BNY Mellon** - *Pune, Maharashtra*   Jun 2022 – Aug 2024

As a Senior Data Scientist with 2.3 years of experience in a leading BFSI organization, I specialize in data science, machine learning, deep learning, NLP, and cloud technologies, contributing to the company's growth and innovation.

- Developed a real-time financial health assessment and credit rating system, increase in process and resource efficiency by 31% and accelerating credit assessment and credit risk evaluation.
- Led the creation of a creditworthiness machine learning solution with a 94% accuracy, reducing maintenance costs and enhancing portfolio management effectively that bring strategizing investment and options planning for investors.
- Employed agile methodologies to cut project delivery time by 20% and streamlining the development of machine learning models and achieving a 40% gain in efficiency.
- Designed and implemented data science and machine learning pipelines for instant financial risk assessment, including developing and validating models for creditworthiness prediction and portfolio management.
- Optimized data science and machine learning with data collection, data preprocessing, data visualization, planning and building ML and deep learning solutions to enhancing both efficiency and accuracy.
- Translated complex data science solutions into actionable insights through collaboration with business stakeholders.
- Managed and maintained data science infrastructure and tools, documenting findings and presenting results to both technical and non-technical audiences.

❖ **Data Administrator - Wipro Technologies** - *Pune, Maharashtra*   Jul 2019 – May 2021

- Client based project working on querying and helping data related issue for the client and partners so make process smooth achieved a client appreciation for working around and provide a satisfactory solution and makes process smoother and increase the productivity.
- Managed and maintained client databases, ensuring accurate and accessible data for internal systems, established secure data backup and recovery protocols for business continuity.
- Streamlined data access through role-based permissions, enhancing security and compliance. Worked on Microsoft Active directory, Microsoft SQL Server, Oracle Identity Management, SNOW for security protocol.
- Provided technical support to internal users regarding data access and troubleshooting database issues.
- Administered SQL Server databases, optimizing performance and conducting routine maintenance.
- Collaborated with cross-functional teams, communicated technical information clearly, and documented procedures for knowledge transfer.

❖ **Trainee – Datamatics Business Solutions** - Nashik*, Maharashtra*   Nov 2018 – Apr 2019

- Experienced with tax preparation software (TurboTax, H&R Block) to streamline tax filing processes.
- Skilled in analysing financial data to identify trends and optimize tax deductions.
- Strong verbal and written communication abilities for effective client and team interactions.
- Proven ability to resolve tax compliance and planning issues efficiently.

EDUCATION

- **Savitri Bai Phule, Pune University** - *Nashik, Maharashtra*   2017 – 2019
  Master of Business Administration [Finance]   8.0 GPA

- **Sant Gadge Baba, Amravati University** - *Yavatmal, Maharashtra*   2014 – 2017
  Bachelor of Computer Application [Computer Science & Engineering]   60 %

SKILLS

- Programming Languages: Python, SQL, Linux, Dataiku, Machine Learning, Deep Learning, AWS
- Data Science: Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, Sklearn, SciPy, Flask, RegEx, win32, openpyxl, PyTesseract, Selenium, PyPDFminur, Smtplib, Pyspark, OS, Boto, OpenCV, OpenAI, TensorFlow – Keras.
- Machine Learning: Linear Regression, Logistic Regression, Naive Bayes, K-NN, Support Vector Machine, K-Means Clustering, Principal Component Analysis (PCA), Decision Tree, Random Forest, Ensemble Learning - Boosting & Bagging Techniques, Ridge & Lasso Regression, Polynomial Regression, Reinforcement Learning.
- Deep Learning: Neural Networks, ANN, CNN, DNN, RNN, LSTM, BERT, Prompt Engineering.
- Data Processing: NLTK, Natural Language Processing, TF-IDF, Word2Vec, Ngrams, Tokenizers, Stemming & Lemmatization,
- AWS: EC2, S3, Sagemaker classic, Lambda, Batch-Processing.
- Tools & Technologies: FastAPI, VS Code, Jupyter Notebook, Git/GitHub, Tableau, Statistics, Data Wrangling, Data Ingestion, Data Mining, Data Modeling & Transformation, Data Analysis, Data Visualization.
- Proficient Verbal and Written Communication, Strategic Issue Resolution, Proven Commitment to Delivering Success, Contribute and Cooperative Approach.

PROJECTS

➢ Banking Fraud Transaction Detection System

Objective: Build a machine learning model to identify fraudulent banking transactions based on predefined rules.
Goal: Predict whether a transaction requires screening for potential fraud.
Approach and Implementation:
- Created rule-based validation features (e.g., transaction amount, Time, and Account type) to flag suspicious transactions.
- Pre-processed data and used machine learning algorithms like Random Forest and XGBoost for classification.
- Evaluated model performance using accuracy, precision, recall, and ROC-AUC.
- Tuned hyperparameters to enhance model accuracy as well implement the Multilayered perceptron.
- Building AI solution to onboard customer to complete the KYC and based on this they also take needful support about product for real-time inference.

Skills: Machine learning, classification, hyperparameter tuning, and fraud detection, Scikit-learn, Pandas, NumPy, TensorFlow.
Impact: Enabled real-time detection of fraudulent transactions, reducing false positives and improving risk management.

➢ Defect Management System for JIRA tickets

Objective: Develop a chatbot-based system to summarize and cluster JIRA tickets for efficient defect management.

Goal: Automatically group and summarize JIRA tickets to improve defect management and tracking.

Approach and Implementation:
- Utilized TF-IDF and cosine similarity to generate summaries of JIRA tickets.
- Applied DBSCAN for defect clustering based on ticket similarity.
- Future Enhancements: Exploring vector-based models such as word2vec and GloVe to enhance clustering accuracy.
- Integrated batch processing to handle bulk tickets, with chatbot interaction for summarization and clustering.
- Generated coloured Excel files to present clustered defects for better clarity and understanding.

Skills:
Expertise in text similarity (TF-IDF, UMAP, Word2Vec, Euclidean distance, cosine similarity) and clustering (DBSCAN), Proficiency with vector-based models (word2vec, GloVe) for advanced NLP.
Automation using Python, batch file processing, and Excel generation.

Impact:
Improved defect tracking and management by automating ticket clustering and summarization, resulting in clearer insights and better resolution processes.

➢ Expleo.ai Low Code No Code Data Platform

Objective: Build a low-code/no-code platform offering automated solutions for data-driven projects across various domains.

Goal: Enable users to accelerate data workflows and deploy machine learning models without coding expertise.

Approach and Implementation:

- Data Acceleration & Analysis: Provided automated data ingestion, preprocessing, and analysis features.

- Visualization & Modelling: Integrated tools for dynamic data visualization, modelling, and preprocessing.

- Machine Learning & Image/Audio Processing: Supported end-to-end machine learning workflows, including image extraction, audio detection, and processing.

- Single-Window Interface: Designed an intuitive interface for domain-specific Proof of Concept (POC) developments, offering seamless integration across tasks.

Skills:

Proficiency in low-code/no-code platforms, data pipelines, and automation, Experience with machine learning, image extraction, audio detection, and processing, Strong knowledge of data analysis, visualization, and model deployment.

Impact:

Enabled rapid prototyping and POCs for diverse domains, accelerating time-to-market for AI and data-driven solutions.

➢ **Credit Risk Portfolio Management with Financial Health – BNY Mellon**

Objective: Diagnose the financial health of businesses and build a machine learning framework to identify companies at risk of defaulting in the next financial year.

Goal: Construct a classification model to assign a probability of default.

Definition of Default: Companies with negative net worth for next year.

Approach and Implementation:
- Utilized various machine learning and deep learning algorithms to develop a classification model for risk assessment.
- Evaluated model performance metrics including accuracy, precision, recall, and ROC-AUC score.
- Applied hyperparameter tuning to optimize model parameters for better performance.
- Employed techniques such as random under sampling, SMOTE, and ADASYN to handle class imbalances.
- Implemented feature scaling and advanced preprocessing methods to enhance model robustness and effectiveness.
- Deployed the model for real-time inference using Amazon SageMaker, Lambda, API Gateway ensuring scalability and reliability.
- Explored deep learning models to assess and compare their capability against traditional machine learning models.

Skills:
- Proficiency in machine learning, deep learning, classification modelling, hyperparameter tuning, and feature scaling.
- Experience with Python libraries including Scikit-learn, Pandas, NumPy, TensorFlow, and Keras for data preprocessing, model training, and evaluation.
- Expertise in deploying machine learning models using Amazon SageMaker.
- Problem-solving and decision-making in financial risk assessment contexts.

Impact:

The project significantly contributed to the development of a reliable tool for financial risk assessment, enabling proactive identification, mitigation, and management of potential default risks.

➢ **Financial Credit Rating Assessment for Individuals – BNY Mellon**

Objective: To optimize financial services for accessing asset and infrastructure management of businesses, thereby facilitating better future investment decisions.

Goal: Develop a machine learning model to accurately classify businesses and assign appropriate credit ratings within a vast financial services portfolio.

Definition of Default: business with individual rating to identify are Long-term Debt or Short-term Debt

Approach and Implementation:
- Contribute the development of robust classification machine learning algorithms for credit rating assessment.
- Integrated diverse financial and non-financial data from income statements, equity and debt report, financial flow to future scope.
- Leverage cloud service SageMaker to build, train, and validate the model.
- Employed techniques such as Logistic Regression and Random Forest Classifier for Credit rating assessment.
- Utilise the Feature engineering and feature scaling to enhance the model efficiency and robustness to make better generalization.

Skills:
- Proficiency in machine learning, classification modelling, hyperparameter tuning, and feature scaling.
- Experience with Python libraries including Scikit-learn, Pandas, and NumPy for data preprocessing and model evaluation.
- Problem-solving and decision-making in financial risk assessment contexts.

Impact: Streamlined the Credit rating and reduce the operation efforts by 31% and with that empowered the informed decision-making across the investors, Collaboration with domain experts and ensure actionable insights for better investment service.

➢ **Ratio Modelling Based on UK - Market Interest Rates Trend – BNY Mellon**

Objective: The challenge is to analyse the dynamics of UK market interest benchmark rates and their relationships with economic factors, aiming to provide insights for financial decision-making.

Goal: To develop regression models that accurately predict interest rate trends based on historical data and economic indicators. The goal is to understand the drivers of interest rates and their impact on financial markets, enabling informed decision-making by stakeholders.

Definition of Default: Index Interest rate trend overnight change in UK markets.

Approach and Implementation:

- Data Collection: Gathered historical data also scrap and align to continue the data processing and along with relevant economic indicators.
- Data Preprocessing: Handled missing values, outliers, and performed feature engineering.
- Model Selection: Utilised appropriate regression techniques such as linear regression, polynomial regression or time series forecasting also used hyperparameter to optimised the results.
- Model Training and Evaluation: Split the dataset into training and testing sets, trained the model, and evaluated its performance using metrics like R-squared and Mean Squared Error.
- Analysis and Interpretation: Analysed coefficients and interpreted results to understand the relationship between economic factors and interest rate trends.

Impact:

The project resulted in an enhanced understanding of rates dynamics for stakeholders, informing their financial decision-making and risk management practices. Additionally, it contributed to the improvement of investment strategies and portfolio management, facilitating policy formulation and market forecasting.