NIKHIL SOMNATH BORADE

Pune, Maharashtra, India

+91 94215 12632 | nikhilsborade3@gmail.com | LinkedIn | GitHub | Portfolio

PROFESSIONAL SUMMARY

AI/ML Engineer with hands-on experience in building, evaluating, and deploying end-to-end machine learning and NLP solutions for classification, regression, and predictive analytics. Skilled in Python, data preprocessing, feature engineering, and model tuning using Scikit-learn and cross-validation. Experienced in deploying production-ready ML applications using Streamlit to solve real-world business problems.

TECHNICAL SKILLS

Languages	Python (primary), SQL (MySQL)
AI / ML	Supervised Learning, Classi cation, Regression, Model Training & Evaluation, Hyperparameter Tuning, Cross-Validation, ML Pipelines, ColumnTransformer
Libraries	NumPy, Pandas, Scikit-learn (KNN, Naive Bayes, Random Forest, Logistic Regression, SVM), Matplotlib, Seaborn, NLTK
NLP	Text Preprocessing, Tokenization, Lemmatization, TF-IDF Vectorization, Text Classi cation, Sentiment Analysis
Data Analysis	Data Cleaning, Exploratory Data Analysis (EDA), Statistical Analysis, Feature Engineering, Correlation Analysis
Deployment	Streamlit (web apps), Pickle (model serialisation), ML Pipelines
Tools	Git, GitHub, Power BI, Microsoft Excel

PROJECTS

Sentiment Analysis using NLP – (AWS App) | | Live Demo — Jan 2026

Python* | *NLTK* | *TF-IDF* | *Scikit-learn* | *Streamlit* | *AWS

- **Problem:** E-commerce platforms generate large volumes of unstructured customer reviews, making manual sentiment analysis infeasible for timely decision-making
- **Approach:** Built an end-to-end NLP pipeline including text cleaning, tokenization, lemmatization, stopword removal, and TF-IDF vectorization; trained and evaluated multiple classi ers such as Logistic Regression, Naive Bayes, Linear SVM, and Random Forest
- **Insight:** Linear models (SVM, Logistic Regression) consistently outperformed ensemble methods on sparse text data, con rming their suitability for high-dimensional NLP tasks
- **Deployment & Result:** Achieved over 92% F1-score and deployed the model as a real-time sentiment prediction web application using Streamlit, hosted on AWS for scalable and production-ready access

Travel Package Purchase Prediction – (Streamlit App) | — Dec 2025

Python* | *Random Forest* | *Streamlit* | *Feature Engineering* | *Scikit-learn

- **Problem:** Sales teams lacked a data-driven method to identify customers most likely to purchase travel packages, leading to inefficient outreach and lower conversion rates
- **Approach:** Performed feature engineering to extract high-signal variables, trained and tuned a Random Forest classi er using cross-validation, evaluated with accuracy, precision, recall, and F1-score
- **Insight:** Customer income, number of follow-ups, and pitch satisfaction score emerged as the strongest predictors of purchase intent
- **Result:** Built and deployed an interactive Streamlit web application that provides real-time purchase probability predictions from customer input
- **Impact:** Enables sales teams to programmatically prioritise high-probability leads, directly improving outreach e ciency and conversion rates

Telco Customer Churn Prediction | — Nov 2025

Python* | *KNN* | *Cross-Validation* | *ROC-AUC* | *Scikit-learn

- **Problem:** Telecom providers lose significant customer base each quarter with no reliable early warning mechanism
- **Approach:** Preprocessed dataset with missing-value handling and StandardScaler; implemented K-Nearest Neighbours with systematic hyperparameter tuning via strati ed cross-validation
- **Insight:** Customer usage frequency, contract type, and service tenure emerged as the strongest behavioural signals predicting churn risk
- **Result:** Achieved 77.7% accuracy and ROC-AUC of 0.73, successfully identifying at-risk customers
- **Impact:** Provides retention teams with ranked at-risk customers for targeted intervention, enabling proactive retention strategies

Fake News Detection using NLP | — Oct 2025

Python* | *NLTK* | *Scikit-learn* | *TF-IDF* | *Naive Bayes

- **Problem:** Misinformation spreads rapidly across online platforms and is extremely di cult to identify manually at scale
- **Approach:** Designed an NLP pipeline with text cleaning, tokenization, stopword removal via NLTK, TF-IDF vectorization, and Multinomial Naive Bayes classi cation
- **Insight:** Recurring word patterns and high-frequency term distributions serve as reliable discriminators between fake and real news articles
- **Result:** Achieved 93.7% classi cation accuracy, demonstrating strong model performance
- **Impact:** Demonstrates the viability of automated, scalable NLP-based content moderation systems for combating misinformation

EDUCATION

Bachelor of Computer Science — July 2022 May 2025

Savitribai Phule Pune University, — Nashik, Maharashtra