

# SHIVAM JOSHI

Jersey City, NJ 07306 | [joshishivam047@gmail.com](mailto:joshishivam047@gmail.com) | (551) 260-8555 | [LinkedIn](#) | [GitHub](#)

## EDUCATION

**Pace University, Seidenberg School of Computer Science and Inform**  
MS in Data Science | **GPA: 3.80**

**New York, NY**  
May 2025

**Mumbai University, BSE Institute Ltd**  
Post Graduate Diploma in Predictive Analytics (PGDPA) | **GPA: 4.00**

**Mumbai, IN**  
June 2023

## TECHNICAL SKILLS

**Programming Languages:** Python (NumPy, Pandas, Scipy), R, SQL  
**Data Science / Machine Learning / NLP:** Scikit-learn, TensorFlow, Keras, XGBoost, NLTK, Spacy, OpenCV  
**Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn, Plotly, Dash  
**Data Mining / Big Data /Cloud:** Hadoop, Spark (PySpark), Google Cloud Platform (GCP), AWS (S3, EC2, Lambda)  
**Databases:** MySQL, MongoDB, PostgreSQL, Oracle, Firebase  
**Tools /Platforms:** Git, Docker, Jupyter, VS Code  
**Certifications:** IBM Data Science (Coursera), Financial Markets by Yale University (Coursera)

## PROFESSIONAL EXPERIENCE

### 99 YARDS

**New York, NY**

Data Science Intern

Dec 2024 – Present

- Developed and deployed a Flask-based recommendation engine using TF-IDF, Firestore queries, and geolocation filtering, improving user engagement by 20% through personalized suggestions.
- Enhanced geospatial intelligence by integrating geopy to prioritize recommendations based on ZIP code proximity, reducing irrelevant matches by 35%.
- Designed a multi-tiered offer categorization logic using price banding, improving usability and clarity for B2B decision making.

### SCIENOX TECHNOLOGIES

**Mumbai, IN**

Data Science Intern

Jan 2023 – Feb 2023

- Conducted time series analysis on 500,000+ stock market records using Python (Pandas, Statsmodels), uncovering key trends and improving forecasting accuracy by 20%.
- Engineered features and applied regression and classification models to enhance model stability, increasing reliability by 18% in investment risk analysis.

### PIPARA & CO.LLP

**Mumbai, IN**

Data Analyst Intern

Jan 2022 – Mar 2022

- Optimized data management processes for annual revenues exceeding INR 5000 Cr, by implementing an analytics framework under IBC 2016, reducing decision-making time by 15%. Identified compliance gaps and improved reporting workflows by 30%.

### M/s UTANGALE & CO. ADVOCATES & SOLICITOR

**Mumbai, IN**

Data Analyst Intern

Oct 2021 – Dec 2021

- Analyzed and maintained financial records for 50+ clients, leveraging data interpretation skills to provide actionable insights and improve client satisfaction.

## PROJECTS

### VinoSage: AI-Driven Wine Recommendation Chatbot ([VinoSage](#))

- Built an AI-powered wine recommendation chatbot using XGBoost, Elasticsearch, and FAISS to deliver personalized suggestions from a dataset of 8,000+ entries, enhancing user experience.
- Implemented NLU components with fine-tuned DistilBERT for intent classification and a BERT-based NER pipeline, enabling accurate and context-aware recommendations.

### Automatic License Plate Detection Using YOLOv5

- Developed an automated license plate detection using YOLOv5, training on a custom dataset of 800+ annotated images to optimize model accuracy.
- Preprocessed and segmented images to create bounding boxes and heat maps, fine-tuning YOLOv5 for improved object detection performance.

### Yield Prediction on Agriculture Dataset

- Analyzed agricultural datasets by preprocessing data (e.g., handling null values, plotting feature correlations) using Seaborn, Matplotlib, and Plotly to identify key factors influencing crop production.
- Trained and evaluated Linear Regression and Decision Tree Regressor models, achieving an R2 score of 0.83 for crop yield.

### Fake News Detection Using Scikit Learn

- Built a real-time fake news classifier using CountVectorizer and PassiveAggressiveClassifier, achieving 91.46% accuracy on test data. Integrated NewsAPI for live article scraping and implemented NLP preprocessing for feature extraction and classification.