Shubham Khode

Chicago, IL | (858) 305-7003 | skhode2@uic.edu | linkedin.com/in/shubhamkhode | shubhamkhode.me

EDUCATION

University of Illinois Chicago *August 2021 – December 2022*
Master of Science in Business Analytics (GPA: 4.0/4.0)

Veermata Jijabai Technological Institute (VJTI), Mumbai, India *August 2012 – April 2016*
Bachelor of Technology in Electronics & Telecommunications Engineering (GPA: 7.5/10)

SKILLS

- **Technology:** Python (NumPy, Pandas, Scikit-Learn), R, SQL, C#, Tableau, Hadoop, Spark, AWS (EC2, S3, Beanstalk), Git, Jira
- **Core Competencies:** Data Mining, Machine Learning, Predictive Modeling (Classification & Regression), Clustering, NLP, Text Analytics, Statistical Analysis, Time Series Forecasting, A/B testing, ETL, Agile Methodologies, DevOps

PROFESSIONAL EXPERIENCE

Wipro Limited
A leading software consulting company; involved with one of the largest telecom operators in Austria (~8M customers) as a client

Business Analyst | Vienna, Austria *August 2019 – June 2021*

- Conducted statistical analysis using SQL and R on customer's products, services, and usage data to discover underlying patterns for mitigating churn rate; made recommendations to bump up retention by 15%
- Designed ETL pipelines to fetch and manipulate data from disparate transaction database systems and consolidate it into data warehouse for downstream BI reporting & analytics
- Applied A/B testing on different landing/product page variants of client's online shop; lifted conversion rate by 7% and marked increase of ~$5M in revenue flow
- Identified borderline credit risk customers using ML algorithms; redesigned order workflows resulting in 2x faster lead time
- Built interactive Tableau dashboards for quick, actionable insights on core business KPIs; saved 8 hrs./week in manual reports
- Collaborated with key stakeholders and leveraged data-driven opportunities for aligning product roadmap to strategic goals

Software Engineer | Pune, India *August 2016 – July 2019*

- Analyzed business requirements, documented high-level technical specifications, and developed E2E solutions ensuring feasibility & integration with legacy systems
- Developed and enhanced 5+ business-critical web applications in .NET to provide a unified experience across all platforms
- Automated log management using Python and prepared dashboards to monitor application activity and performance; improved MTTR (Mean Time to Resolve) for production bugs by 35%, saving ~$500K/year
- Implemented complex SQL scripts for applying pseudonymization; transformed 1M+ customer records with sensitive information using data masking procedures for EU GDPR compliance

PROJECTS

Forecasting Retail Sales for Walmart at scale *(Python, Apache Spark, Facebook Prophet)*

- Analyzed influence of factors like markdowns, holidays, fuel price, temperature, etc. on weekly sales over 450K+ datapoints.
- Trained hundreds of time series models in parallel by leveraging Spark distributed data processing and generated precise forecasts for each store-department combination using Prophet library.

Loan Default Prediction and Investment Decisions for Lending Club *(R, EDA, Classification, Regression)*

- Performed EDA (univariate, multivariate, correlation) to understand loan characteristics and driving features behind defaults.
- Built Decision Tree, Random Forest, and GLM models to predict the probability of loan default and estimate expected annual returns. Evaluated performance using AUC, ROC, and lift curves and combined best models that maximized ROI.

Spotify – Top Songs & Artists Popularity and Recommendation *(Python, Tableau)*

- Investigated and visualized various acoustic features to answer riveting questions pertaining to song and artist popularity.
- Experimented with different regressors using GridSearchCV to predict popularity of song. Designed a recommendation system based on neighborhood collaborative filtering with Cosine and Manhattan distances as similarity metrics.

Market Segmentation of Soap Company *(R, Clustering)*

- Executed K-means, K-medoids, and Hierarchical clustering algorithms on purchase behavior data. Profiled customers based on clusters identified using Elbow & Silhouette methods and proposed strategies that targeted individuals in each segment.

Text Sentiment Analysis on Yelp Reviews *(R, NLP)*

- Applied TFIDF text vectorization on restaurant reviews to classify sentiment polarity (positive/negative) using AFFIN, Bing, and NRC dictionaries. Logistic Regression with LASSO regularization performed best with an overall F1 score of 0.92.