

Bhaskar Shivaji Kumbhar

Tempe, AZ | bkumbhar@asu.edu | (623) 340 - 1336 | Youtube | LinkedIn | Github

Objective

I am a Master's student in Data Science, Analytics, and Engineering at Arizona State University, working toward becoming a data scientist with a strong focus on artificial intelligence and machine learning. My plan includes completing a detailed Master's thesis and continuing my studies in a PhD program. I am passionate about using these opportunities to advance my skills in technology and research. Eventually, I aspire to become a professor to share my knowledge and experience, helping to shape future experts in the field. My goal is to connect academic research with real-world applications to innovate and improve the areas of data science and technology.

Education

Arizona State University, MS: Data Science, Analytics and Engineering August 2024-May 2026

- GPA: 3.89/4.0
- **Key Coursework:** Delving into Statistical Machine Learning (CSE 575) and Data Mining (CSE 572), I have developed strong analytical skills and a deep understanding of modeling techniques which are crucial for tackling complex data-driven challenges in various industries. Additional coursework includes Data Processing at Scale (CSE 511), Statistics for Data Analysts (DSE 501), Computing for Data-Driven Optimization (DSE 506), and Data Science for System Decision Analytics (IEE 577).

University of Mumbai, BE: Information Technology August 2018-May 2022

- CGPA: 8.91/10.0
- **Key Coursework:** Focused on Python, Artificial Intelligence, and Machine Learning, complemented by studies in Database Systems, Business Analytics, and both Web and Mobile Application Development. My comprehensive foundation also includes significant training in Operating Systems, Computation Theory, and robust programming skills in Java, C, and R.

Professional Experience

Software Engineer, LTIMindtree, Navi Mumbai, India June 2022-August 2023

- Successfully completed a comprehensive one-month training program that covered Java, Spring Boot, Hibernate, and Angular. This rigorous training culminated in the design and implementation of a sophisticated credit management system.
- Led a small, dedicated team to develop a robust multi-phase system that enhanced UI improvements and backend integration using Angular and Spring Boot.
- Implemented complex CRUD operations using Kafka and ActiveMQ, establishing secure and efficient data management protocols that significantly improved system functionality and security.
- Achieved multiple industry-recognized certifications including JAVA SE 11 OCP, and Microsoft certifications in Azure, AI, Security Fundamentals, and Dynamics 365.
- Enhanced professional skills communication and adapted to U.S. cultural practices, mastered key compliance protocols including Anti-Bribery, Corruption Protocols, Data Privacy, and Information Security, and actively contributed to the business and finance (BFA) department's objectives by leading strategic internal projects that drove significant operational improvements.

Project Intern, Tata Institute of Fundamental Research, Mumbai, India August 2021-April 2022

- Led the development of a Package Management System at **TIFR**, a leading research institution under the **Department of Atomic Energy of the Government of India**, creating a database-driven web application that streamlined package handling and management, supporting over 500 weekly user interactions without downtime.
- Deployed the system on TIFR servers, integrating with official website templates to enhance UI and backend functionalities, transitioning from a manual to a fully digital system.
- Developed a QR code-based Android application to modernize parcel notifications and collection processes,

significantly reducing paperwork by 90% and enhancing operational efficiency.

- Implemented automated email reminders for uncollected parcels, improving administrative efficiency and reducing parcel backlog.
- Adhered to strict TIFR security protocols, ensuring all developments complied with government standards for information security and privacy.

Publications

Mental Health Issues Prediction during Covid Pandemic Using Machine Learning

April 2022

Bhaskar Kumbhar, Khadija Morbiwala, Snehal Phapale, Radiowalla Ateef Imran

International Journal of Innovative Research in Engineering

Volume 3, Issue 2 (March-April 2022), PP: 124-127.

Abstract : This study assesses the impact of the COVID-19 pandemic on student mental health, using a dataset focused on social life and mental health factors. We developed a web application that utilizes a machine learning model to predict mental health status based on user inputs. The model employs a Decision Tree classification algorithm, chosen for its accuracy in predicting various mental health issues based on symptoms like anxiety, depression, and stress. Our findings indicate significant mental health challenges among students during the pandemic, highlighting the utility of our model in providing timely interventions.

Key Contributions:

- **Data Analysis:** Analyzed a dataset from Kaggle, consisting of students' lifestyle data during the pandemic to identify key predictors of mental health issues.
- **Model Development:** Implemented and compared several machine learning algorithms, ultimately selecting the Decision Tree for its superior performance in our context.
- **Application Development:** Created a user-friendly web application that allows students to assess their mental health status, offering tailored recommendations based on the model's predictions.
- **Impact:** The system aims to raise awareness and provide early diagnostic support, potentially guiding interventions to assist students struggling with pandemic-related mental health issues.

Technical Skills

Languages: Python, R, Java, C, C++ , PHP, SQL

Database: MySQL, SQLite, PostgreSQL, MongoDB

Tools and Platforms: Jupyter Notebook, Google Colab, VS Code, PyCharm, Git, Docker, Kubernetes, PowerBI, Tableau, Excel, Apache Spark, Hadoop

Machine Learning Libraries: Scikit-Learn, TensorFlow, PyTorch, Keras, XGBoost, LightGBM, CatBoost, H2O.ai

Deep Learning and AI Concepts: Neural Networks (CNN, RNN, LSTM, GANs), Transfer Learning, Reinforcement Learning, Autoencoders, Attention Mechanisms, Transformer Models

Natural Language Processing: NLTK, spaCy, Hugging Face Transformers, BERT, GPT, Word2Vec, GloVe, TextBlob

Computer Vision: OpenCV, FastAI, YOLO, TensorFlow Object Detection API

Big Data and Cloud Platforms: AWS (EC2, S3, Lambda), Microsoft Azure, Google Cloud Platform (GCP), Apache Spark, Databricks

Hyperparameter Tuning Libraries: Optuna, Hyperopt, GridSearchCV, RandomizedSearchCV

Deployment Tools: Heroku, Streamlit

Operating Systems: Windows, Linux, Ubuntu, Unix

Other Technologies and Concepts: Data Structures and Algorithms, Statistics, Probability, Time Series Analysis, Data Mining, Data Visualization, Feature Engineering, Model Deployment (Flask, FastAPI), MLOps, Agile Methodology, Spring Boot, REST APIs

Projects

CineGraph Cypher AI: Intelligent Movie Insights with Neo4j Groq LLM

March 2025

- Developed a Graph-Based Movie Database using Neo4j, enabling structured storage and efficient querying of movie relationships, including actors, directors, genres, and IMDb ratings.
- Integrated Groq's LLaMA 3.3-70B LLM with Neo4j to allow natural language queries on movie data, achieving over 90% accuracy in retrieving movie details such as directors, genres, and IMDb ratings.
- Enhanced query execution speed, improving movie search performance by 45% over relational databases through optimized Cypher queries.
- Implemented intelligent recommendations by leveraging LLM-generated responses, enabling context-aware movie exploration.
- Built a scalable and efficient graph schema, ensuring fast retrieval of highly connected movie data.
- Bridged structured data storage with AI-powered search, demonstrating expertise in Graph Databases, NLP, and intelligent information retrieval.

Drought Prediction : Enhancing Forecasting Accuracy and Resilience

Oct 2024 - Dec 2024

- Led the development of a machine learning project utilizing a 20-year dataset from the NASA POWER Project and U.S. Drought Monitor, encompassing over 1 million records to enhance drought prediction models.
- Utilized advanced data mining and machine learning techniques, specifically GPU-accelerated XGBoost, to process and analyze the extensive dataset, achieving a remarkable R^2 score of 0.828 on validation and 0.826 on test data, thereby demonstrating the model's robust predictive capability compared to other algorithms like CatBoost and LightGBM.
- Addressed data imbalances and integrated complex temporal features, improving drought classification accuracy by 15% through advanced feature engineering.
- Conducted rigorous model evaluations with cross-validation, ensuring reliability and accuracy across varied conditions.
- The project's findings offer valuable insights for improving drought management strategies, aiming to mitigate the impacts of droughts on communities and economies.

Comprehensive Analysis of Arizona Businesses using the Yelp Dataset

Sept 2024 - Dec 2024

- Analyzed over 10 million reviews and user interactions from the Yelp dataset using Apache Spark and SQL, deriving insights into restaurant trends and how amenities like WiFi and outdoor seating influence customer ratings.
- Identified that amenities such as free WiFi and outdoor seating contributed to a rating increase of up to 3%, demonstrating their value in enhancing customer satisfaction and attracting more patrons.
- Performed a user-level analysis of over 100,000 users, revealing that elite users with over 100 friends play a crucial role in shaping review sentiments, generally giving higher ratings (average 3.90) and producing more detailed reviews.
- The project's findings provided actionable insights that helped restaurant owners and managers in Arizona tailor their services to better meet consumer expectations, leading to improved customer retention and business success.

Multimodal Text Analytics Framework for Semantic Classification and Predictive Modeling

Aug 2024 - Dec 2024

- Developed a Python-based text classification system using TF-IDF, CountVectorizer with N-Grams, and GloVe embeddings, processing over 1,000 news articles to enhance feature extraction and text understanding.
- Achieved a 95.3% accuracy in semantic classification by implementing Random Forest and MLP classifiers, evaluated through 5-fold stratified K-fold cross-validation.
- Enhanced prediction reliability by 20% through comprehensive hyperparameter tuning and streamlined data processing.
- Implemented automation for structured CSV output generation, enhancing scalability and facilitating broader data accessibility for text analytics.

Intrusion Detection System Using Statistical Machine Learning

Sept 2024 - Nov 2024

- Led the development of an Intrusion Detection System (IDS) designed to detect and classify network intrusions, focusing on Distributed Denial of Service (DDoS) and PortScan attacks, utilizing a comprehensive dataset from the Canadian Institute for Cybersecurity.
- Engineered the system using a variety of machine learning algorithms including Simple Neural Networks (SimpleNN), Deep Neural Networks (DNN), 1D Convolutional Neural Networks (1D CNN), and a hybrid model combining Autoencoders with Random Forests, to effectively recognize and analyze complex patterns and anomalies in network traffic.
- Conducted rigorous data preprocessing which included normalization, handling missing values, and strategic dataset splitting into training, validation, and testing subsets to enhance the training process and model evaluation.
- Achieved exemplary performance metrics, with the hybrid Autoencoder and Random Forest model reaching an accuracy of 99.91%, demonstrating superior predictive capability and reliability over other models like DNN, which also performed well with an accuracy of 99.59%.
- Demonstrated the effectiveness of machine learning in enhancing traditional intrusion detection capabilities, providing actionable insights and methodologies that significantly advance network security protocols.

Hyperparameter Tuning in Time Series Forecasting

Oct 2024

- Conducted a detailed study on hyperparameter tuning for a time series prediction model using a Multi-Layer Perceptron (MLP), hosted as a competition on Kaggle.
- Achieved significant improvements in model accuracy through meticulous testing and adjustment of hyperparameters, including learning rates and epoch counts.
- Found optimal settings that effectively balanced training efficiency and prediction accuracy, evidenced by a notable decrease in Mean Squared Error (MSE).

IMDB Movie Recommendation System

Mar 2022

- Designed and implemented a movie recommendation system using R, leveraging popular libraries such as recommenderlab and ggplot2 for data analysis and visualization.
- Utilized collaborative filtering algorithms within the recommenderlab package to process IMDB movie ratings data, significantly enhancing the system's ability to predict user preferences.
- Achieved an impressive accuracy rate, with the system successfully predicting user movie preferences with over 85% accuracy, facilitating more personalized and relevant movie suggestions.
- Developed a user interface in R Shiny to allow for easy access and interaction by users, making the recommendation process both engaging and user-friendly.

IMDB Data Analysis Project

Mar 2021

- Conducted a comprehensive analysis of the IMDB dataset to uncover trends and insights into movie ratings, genres, and viewer preferences using Tableau.
- Utilized Tableau's robust data visualization tools to create interactive dashboards and graphs that highlighted key patterns and anomalies in the data, making complex information easily understandable.
- Identified significant trends such as the impact of directors and actors on movie ratings, and the popularity of genres over different periods, providing valuable insights for industry stakeholders.
- Presented findings in a series of well-received reports and visual presentations, enhancing decision-making processes for media and entertainment professionals.

Android Quiz Application - 'KnowledgeBee'

Oct 2021

- Engineered an interactive Android quiz application called 'KnowledgeBee', designed to offer a dynamic learning environment through quizzes across multiple categories including Drama, Science, Mathematics, Puzzle, and General Knowledge.
- Implemented user authentication and account management with Firebase to ensure secure and personalized user access to the application.

- Utilized Firestore for real-time database services to store and fetch quiz questions dynamically, ensuring a varied and unpredictable quiz experience by delivering random questions each session.
- Incorporated gamification elements such as a scoring system where correct answers reward users with coins, and a spin wheel feature for additional coin opportunities, enhancing user engagement.
- Integrated a real-time leaderboard to foster a competitive environment, displaying rankings based on the coins accumulated by users.
- Designed a user-friendly interface to ensure ease of use and accessibility, making the application suitable for users of all ages and technical abilities.

Online Venue Booking System - 'BookingWala'

Oct 2021

- Developed a multi-role online venue booking system allowing customers to book venues, view historical transactions, and venue owners to manage listings and track profits, all via a responsive web interface.
- Integrated Google DialogFlow to create a chatbot that assists venue owners with real-time inquiries, enhancing user interaction and operational efficiency.
- Developed functionalities for different user roles: customers can book and view venues, venue owners can update venue details and images, and admins can manage transactions and adjust pricing.
- Implemented financial features where admins receive a 10% commission on bookings, and venue owners can view their earnings and booking history, ensuring a comprehensive management platform.
- Focused on creating a user-friendly environment using Bootstrap for easy navigation, ensuring the platform's accessibility to users of varying technical skills.

Fitness and Nutrition Website - 'FitnessForSUB'

Nov 2020

- Developed a comprehensive fitness and nutrition website to provide users with resources on home workouts, nutrition, and meditation, featuring videos, images, and information.
- Implemented secure user authentication features, including login and registration functionalities with password hashing to ensure data security.
- Created interactive tools such as a BMI (Body Mass Index) calculator using JavaScript, allowing users to easily assess their fitness levels directly on the website.
- Designed a contact form using PHP where users can submit queries, enhancing user engagement and support.
- Ensured the website is fully responsive, adapting seamlessly to different device screens, enhancing user experience and accessibility.

Basic Banking System

Mar 2021

- Designed and implemented a secure, responsive web-based banking system that enables users to perform financial transactions and access their transaction history.
- Built using a robust technology stack including HTML, CSS, PHP, JavaScript, and MySQL, ensuring comprehensive functionality.
- Enhanced user experience by employing Bootstrap for a responsive design, adapting seamlessly across different devices.
- Deployed the application on AWS as part of Cloud Computing coursework, demonstrating proficiency in cloud services and deployment.

Volunteering

Data Analytics Mentor

Jun 2023 - Dec 2023

- Mentored commerce and arts background students in Data Analytics, delivering comprehensive training in Python, SQL, Excel, Tableau, and PowerBI.
- Designed and conducted hands-on sessions focusing on real-world data challenges, enabling students from non-technical backgrounds to grasp and apply data analytics concepts effectively.
- Guided students through practical projects and case studies, enhancing their analytical skills and preparing them for industry roles in data-driven fields.

Social Team Member, Association for Computing Machinery (ACM)

Aug 2019 - May 2021

- Collaborated with team members to organize fundraising campaigns during undergraduate studies, successfully collecting funds, clothes, books, and toys for donation to Bal Asha Trust (Orphanage) and an Old Age Home.
- Actively participated in outreach activities, engaging with community members to promote social causes and encourage participation in charitable events.
- Contributed to planning and executing donation drives, ensuring the collected resources were efficiently distributed to support underprivileged children and elderly residents.