

PRAGATI KHEKALE

773-998-2212 | khekalepragati01@gmail.com | IL, US | <https://www.linkedin.com/in/pragati-khekale/> | <https://pragati2.github.io/PragatiKhekale.github.io/>

SUMMARY

I am a recent data science graduate, with work experience as a data scientist and Data Analyst which enhanced sk pills to convert data into insights for driving business decisions. Experienced in Machine Learning, Data Analytics, Data Visualization, and Predictive Analytics.

EDUCATION

Illinois Institute of Technology, Chicago, USA.

Jan 2021 – Dec 2022

Master of Science, Data Science (Major: Computer Science & Statistics)

Savitribai Phule Pune University, Pune, India

Aug 2016 – May 2020

Bachelor of Engineering, Electronics and Telecommunication

SKILLS

Programming Languages: Python, R, C, C++

Analytical Tools: Tableau, MS Excel, Power BI, Hadoop (Hive, Impala, Spark), Google Analytics, Statistical Analytics

Other Tools & Cloud: Git, Linux, LaTeX, Bitbucket, Trello (Project Management), AWS, Google Cloud, UNIX

Database Systems: MySQL, MS SQL server, PostgreSQL

Functional Skills: Database Management, Communication, Risk Analytics, Agile, Leadership, Data Analysis

Other Technical Skills: NumPy, OpenCV, Matplotlib, Pandas, Seaborn, TensorFlow, Keras, Scikit-Learn, Hypothesis Testing, PyTorch

PROFESSIONAL EXPERIENCE

Data Scientist Intern | Nebal Technologies | IL, US

Sep 2022 – Present

- Visualize, interpret, report findings, and develop strategic uses of data by python Libraries like Numpy, Scikit-learn, Matplotlib and seaborn
- Implemented deep learning algorithms such as Artificial Neural network (ANN) and Recurrent Neural Network (RNN), tuned hyper-parameter and improved models with Python packages TensorFlow.
- Cleaned data using R, then visualize the data, and derive statistical modelling plots.
- Created multiple custom SQL queries in Teradata SQL Workbench to prepare the right data sets for Tableau dashboards. Queries involved retrieving data from multiple tables using various join conditions that enabled to utilize efficiently optimized data extracts for Tableau workbooks.
- Monitoring and tracking process performance using analytics tools like Tableau dashboard, R.
- Participated in all phases of data mining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis.
- Designed and developed new reports and maintained existing reports using Microsoft SQL Reporting Services (SSRS) and Microsoft Excel to support the firm's strategy and management

Data Scientist | CCC Intelligent Solutions | Chicago, US

May 2022 – Aug 2022

- Build predictive models using Machine Learning techniques & evaluated model effectiveness to generate insightful analytics for improved accuracy & reliability in business models.
- Engineered SISA architecture to examine 250,000 data points through 6 scenarios for model evaluation & assessment.
- Analyzed comparative performance of different hardware shards for exploration and optimization of SISA model in real-time.
- Facilitated comparative analysis for 250K data points across 6 scenarios, yielding real-time performances of the SISA model along various shards with average accuracy of 96%.

Data Analyst | Trigent | India

Oct 2019 – Dec 2020

- Analyzed different types of data to derive insights about relationships between locations, statistical measurements and qualitatively assess the data using R
- Used SAS for pre-processing data, SQL queries, data analysis, generating reports, graphics, and statistical analyses.
- Performed data manipulation, analysis, modeling, and visualization on massive transaction data with Impala, Python and R Collaborated with Business.
- Programmed a utility in Python that used multiple packages (SciPy, NumPy, pandas)
- Designed, developed and maintained daily and monthly summary, trending and benchmark reports repository in Tableau Desktop.
- Used SAS/SQL to pull data out from databases and aggregate to provide detailed reporting based on the user requirements.

- Developed Tableau data visualization using Scatter Plots, Geographic Map, Pie Charts and Bar Charts and Density Chart
- Performed data analysis and data profiling using complex SQL on various sources systems including Oracle and Teradata
- Perform data manipulation, development and validation of data models using R and SQL

Responsible for defining the key identifiers for each mapping/interface PROJECTS COMPLETED

- **Oral Cancer Detection Using Deep Learning Application** Aug 2019 – Sept 2020
 - Gathered 6500 X-ray image datasets from multiple sources, labelled & classified data to train supervised model using VGG-16 architecture with 91% accuracy. Trained model resulted in 35% more accurate predictions than existing method while reducing time complexity by 40%.
- **OpenMax Open Set Deep Networks** Jan 2022 – May 2022
 - Constructed OpenMax OSDN model with object detection to increase precision & classifier performance, resulting in a 12% decrease in false positives. Deployed and optimized machine learning models to reduce false positives using OpenMax optimizer for open-set deep networks. Optimized open set deep networks classifier using Acuity Model Compression Module, leading to a 75% reduction in detection time.
- **Quantified Machine Learning Focused Analysis of multiple AWS services.** Jan 2022 – May 2022
 - Built & integrated Logistic Regression, SVM & Gradient Boost Machine Learning models with AWS services to analyze structured and unstructured data with 98% accuracy. Devised scalable Cloud Computing solution to implement and batch process machine learning models leveraging EC2 instances and S3 buckets, resulting in a 70% faster execution time.
- **Black Friday Sales Prediction** Aug 2022 – Dec 2022
 - Constructed database from CSV format with 750K records to enable multivariate analysis of customer's product segment & sales metrics, resulting in an 86% prediction accuracy. Executed Random Forest Regressor, Gradient Boosting Regressor and XGBoost Regressor models to create sales predictions. Deployed single and multivariate analysis on 12 columns of predictor variables, boosting forecast accuracy by 12%.