

Manisha Lagisetty

(925) 922-5765 | manishalagisetty@gmail.com | [Portfolio](#) | [LinkedIn](#) | [GitHub](#)

Education:

Master of Science in Data Science, San Jose State University, San Jose, California Jan 2023 – Dec 2024

- **Coursework:** Machine Learning, Deep Learning, Data Mining, Big Data Technologies, Data Visualization, Database Systems

Bachelor of Computer Science & Engineering, Gitam University, India Jun 2014 – Apr 2018

- **Coursework:** Implemented Authorship Attribution using K-Means Clustering for Capstone Project

Skills:

- **Programming/Scripting:** Python, PySpark, SQL
- **Big Data:** Hadoop, Spark, Kafka
- **Machine Learning Statistics:** Regression, Classification, Time-series analysis, CNNs, GANs, NLP
- **Cloud Platforms and Data Visualization:** Aws, Azure, Tableau
- **Data Engineering:** ETL, Data Pipelines, Data Modeling, Data Integration, Data Migration, Data Lineage, A/B Testing
- **Version Control Systems, Tools and Methodologies:** Git, GitHub, SQL Workbench, Microsoft Office Suite, Trello, Jupyter Notebook, Anaconda, Agile, Waterfall, Scrum, JIRA, Confluence
- **Certifications:** AWS Certified Cloud Practitioner (CLF-C02), Certified Microsoft Azure Fundamentals (AZ-900)

Work Experience:

Data Analyst, Cognizant Technologies Oct 2018 – Jun 2022

- Analyzed intricate healthcare data to develop and implement **A/B testing strategies** that enhanced user interaction and experience with insurance platforms.
- Developed and implemented **SQL**-based business rules, **Python** scripts and **ETL** processes to automate healthcare operations, leading to significant cost reductions and a 20% improvement in processing time. Enhanced productivity by streamlining data workflows and improving data accuracy.
- Created and monitored **KPIs** to assess the performance of healthcare initiatives. Designed and evaluated hypothesis-driven experiments to test and optimize strategies, improving overall business performance. Developed dynamic dashboards that provided real-time insights to stakeholders, enabling data-driven decision-making.
- Collaborated with **cross-functional teams** to enhance data segmentation models and visualizations. Used **data visualization tools** to create automated reports that clearly communicated complex data trends, helping to align business strategies with evolving market conditions.

Projects:

Wine E-Commerce Application Jan 2023 – May 2023

- **Objective:** Analyze marketing trends and revenue to optimize business performance and identify potential issues.
- **Implementation:** Developed an integrated and interactive analytical application using Python (PyQt5) for real-time KPI monitoring and insights into customer behavior. Leveraged MySQL for data storage and retrieval, while utilizing ETL processes with NumPy and Pandas to streamline data extraction, transformation, and loading, enabling efficient data-driven decision-making.
- **Tools:** Python(PyQt5 GUI, NumPy, Pandas), MySQL, SQL Workbench

A Cross Country Region Wise Analysis of Adolescent Delinquency Jan 2023 – May 2023

- **Objective:** To understand the root causes of adolescent delinquency and identify potential interventions.
- **Implementation:** Conducted a comprehensive analysis of large-scale adolescent delinquency datasets, utilizing Tableau for data visualization and statistical analysis to uncover patterns, correlations, and regional hotspots of delinquent behavior. The findings were presented through interactive dashboards, enabling data-driven decisions for effective interventions.
- **Tools:** Tableau, Python(Pandas, NumPy)

Sustainable Future through Natural Disaster Prediction Oct 2023 – Dec 2023

- **Objective:** Enhance global resilience by predicting future natural disasters and providing early warnings for proactive measures.
- **Implementation:** Developed and fine-tuned machine learning models, including classification and time-series analysis, to forecast natural disasters using historical data. Employed Python for data preprocessing and model training, ensuring high accuracy and reliability in predictions to support early warning systems and proactive actions.
- **Tools:** Python (Pandas, Scikit-Learn, Matplotlib, Seaborn), Machine Learning(Classification, Time-Series Analysis)

Predicting Crime and Proposing Safer Neighborhoods

Oct 2023 – Dec 2023

- **Objective:** To create a predictive system for identifying high-risk areas and proposing data-driven strategies for safer neighborhoods.
- **Implementation:** Analyzed historical crime data to uncover patterns and develop predictive models using advanced data analysis techniques. Achieved a 91% F1-Score with tree-based models, enhancing resource allocation and strategic planning for crime prevention.
- **Tools:** *Python (Pandas, Scikit-Learn, Matplotlib, Seaborn), Machine Learning(XGBoost, Random Forest, Decision Tree)*

Travel Recommendation System

Oct 2023 – Dec 2023

- **Objective:** Develop an advanced analytics-driven travel recommendation system that provides personalized suggestions based on individual preferences.
- **Implementation:** Processed and analyzed complex datasets from multiple sources to identify patterns and trends, enabling personalized recommendations using content-based and collaborative filtering techniques. This system improved travel planning by offering precise, tailored suggestions, enhancing user satisfaction through data-driven decisions and optimized performance.
- **Tools:** *Python (Pandas, Matplotlib, Seaborn), Machine Learning*

AI-Driven Application for Diabetes Care Predictive: Analysis and Personalized Recommendations

Jan 2024 – May 2024

- **Objective:** To improve diabetes care by utilizing predictive analytics and offer personalized health recommendations.
- **Implementation:** Designed an intuitive, real-time application that employs machine learning models to assess diabetes risk using individual health data. The system delivers highly accurate predictions and customized advice, enabling users to actively manage their condition with personalized, reliable lifestyle and dietary recommendations.
- **Tools:** *Python (Pandas, Matplotlib, Seaborn, streamlit), Machine Learning and Data Mining techniques*

The Learning Agency Lab: PII-Data-Detection (Kaggle Competition)

Jan 2024 – May 2024

- **Objective:** To develop a model that detects personally identifiable information (PII) in student writing.
- **Implementation:** Applied state-of-the-art Natural Language Processing techniques to analyze and identify PII-related patterns, including names, email addresses, and identification numbers. The model was designed to enhance data privacy by accurately detecting sensitive information in educational texts, improving safety in learning environments.
- **Tools:** *Python, Natural Language Processing (NLP) techniques, Deep Learning(DistilBERT)*