Mallikarjuna Mannem - Data Scientist

616-315-9830 | mannemm@mail.gvsu.edu

EDUCATION

Master of Science in Data Science and Analytics — **May 2025**
Grand Valley State University, Grand Rapids, MI — GPA: 4.0

Bachelor of Technology in Computer Science and Engineering — **April 2018**
Andhra University, Visakhapatnam, India — GPA: 3.7

SKILLS

- Machine learning, statistical computing, modeling & graphics with R, knowledge discovery & data mining, ETL processes, and data pipelines, predictive analysis, exploratory analysis, schema design, data engineering & information management (NoSQL and SQL databases), data visualization and UX design
- R, Python, C, SQL, PHP, HTML, CSS
- MS Excel, Power BI, Git, Azure fundamentals, Ansible, Figma, Miro, Canva, Pentaho, Google Colaboratory.

RELATED PROJECTS

Machine Learning: Unveiling Viewer Engagement through YouTube Video Data Analysis

Objective: Analyzed YouTube video data to understand the relationship between video characteristics and viewer engagement using sentiment analysis of comments.

Data Sources: Video metadata and viewer comments datasets.

Techniques:

- Data Cleaning: Removed non-alphanumeric characters, stop words, and emojis.
- Feature Extraction: Used CountVectorizer and TruncatedSVD.
- Sentiment Analysis: Employed TextBlob for sentiment polarity scores; visualized using violin plots.
- Model Training: Trained Random Forest, Logistic Regression, and Multinomial Naive Bayes models.
- Evaluation: Cross-validation metrics (accuracy, precision, recall) using GridSearchCV.

Key Findings:

- Weak correlation between comment sentiment and video views (Pearson coefficient 0.017).
- Frequent comment terms identified through word cloud analysis.

Tools: Python, Pandas, NLTK, TextBlob, Scikit-learn, Matplotlib

Impact: Provided insights for content creators and platform managers to enhance viewer engagement and content recommendation systems.

Data Mining: Michigan Surface Water Analysis

Objective: Analyzed surface water data in Michigan to study the effects of rising PFAS (Per- and Polyfluoroalkyl Substances) contamination and the Flint water crisis.

Data Cleaning and Preprocessing:

- Handled missing values and inconsistent data formats.
- Converted 'CollectionDate' to numerical features representing days.
- Scaled numerical data for better model performance.

Feature Engineering:

- Engineered features such as 'CollectionDate' to ordinal format for improved model accuracy.
- Combined encoded features with numerical features like 'X' and 'Y' coordinates.

Exploratory Data Analysis:

- Created scatter plots and time series visualizations to illustrate PFAS trends.
- Analyzed PFAS compound concentrations over time and across different locations.

Model Development:

- Developed and evaluated Linear Regression and Random Forest models to predict PFAS concentrations.
- Conducted hyperparameter tuning using Grid Search Cross-Validation to optimize model performance.

Future Work:

- Investigate additional machine learning models for better prediction accuracy.
- Explore more advanced feature engineering techniques to capture complex relationships in data.

Publication Plan:

- Collaborate with Professor Kamrul Hassan to write and publish an article on PFAS contamination and its implications for public health in Michigan.
- Document findings, methodologies, and insights gained from the analysis to contribute to environmental science literature.

Tools and Technologies:

- Utilized Python for data cleaning, preprocessing, and analysis.
- Leveraged libraries such as pandas, scikit-learn, matplotlib, and seaborn for data manipulation, model development, and visualization.

R Programming: Gun Violence Data Analysis

Objective: Analyzed gun violence incidents in the United States to identify patterns and trends in gun-related incidents, including fatalities and injuries.

Data Cleaning and Preprocessing:

- Handled missing values and inconsistent data formats.
- Standardized date formats and extracted relevant date components (e.g., year).
- Merged geographic data with incident data for comprehensive analysis.

Exploratory Data Analysis:

- **Visualizations**: Created scatter plots, histograms, and pie charts to visualize key trends in gun violence incidents.
- **Incident Distribution**: Plotted incidents by year and state to identify temporal and geographic trends.
- **Casualty Analysis**: Analyzed the number of fatalities and injuries to determine the severity of incidents over time.

Participant Gender Analysis:

- Extracted and analyzed participant gender data to understand the demographic distribution of individuals involved in incidents.
- Created a pie chart displaying the proportion of male and female participants.

Geospatial Analysis:

- Mapped the geographic distribution of incidents using latitude and longitude data.
- Generated heatmaps to identify regions with high density of gun violence incidents.

Interactive Visualization:

- Developed an interactive plot to display the location of incidents, distinguishing between killed (red) and injured (yellow) participants, with incident details shown on hover.

Tools and Technologies:

- Utilized R for data cleaning, analysis, and visualization.
- Leveraged libraries such as `dplyr`, `ggplot2`, and `leaflet` for data manipulation and plotting.
- Applied advanced data wrangling techniques to handle complex and nested data structures.

Data Engineering:

- SQL databases like MySQL and Postgre. Scored 'A' for the assignments and the final test.
- Implemented CRUD operations in NoSQL databases like MongoDB, Neo4J, REDIS, and TimescaleDB.
- Used Pentaho data integration tool and performed analytics through ETL process.
- Databases like boats and types, stocks and nations, student profiles, and social media campaigns.

Applied Statistics in Health Professions:

- Exploratory data analysis, categorical predictors and dummy variables, one-sample t-test, two-sample t-test, and means t-test using the Penguins dataset.
- One-way ANOVA (analysis of variance) test using Student GPA and Cancer Survival datasets.
- Fisher exact test using Cookies experimental data.
- Simple linear regression using Scrabble dataset, and Multilinear regression using Great Lakes data.

Sporta - DesignBoat:

- Spearheaded user experience design for a multifaceted sports application encompassing club activities, venue rentals, equipment transactions, and event notifications.
- Conducted comprehensive user research, leveraging interviews, surveys, and observations to glean insights into user preferences and behaviors, and subsequently crafted detailed user personas.
- Employed information architecture tools, including sitemaps and user flows, to meticulously organize content and ensure intuitive navigation.
- Developed low-fidelity wireframes and interactive prototypes, facilitating iterative testing and refinement to enhance usability and functionality.
- Tools used: Figma, Miro, Canva, FigJam.

PROFESSIONAL EXPERIENCE

Data Storage Consultant | CGI, INDIA **Mar 2023 – Aug 2023**

- Upgraded NetApp clusters (200) across multiple customers.
- Executed data transfer projects from different storage boxes to NetApp using Komprise, Robocopy, and Snapmirror. Supported the Keystone environment for NetApp clients.
- Storage technologies: NetApp, Unity, Azure Fundamentals, and related software tools.

Data Storage Administrator | Tata Consultancy Services, INDIA **July 2018 – March 2022**

- Collaborated with cross-functional teams to assess storage requirements and implement scalable solutions that meet business needs.
- Conducted routine maintenance, upgraded storage systems, and troubleshot storage systems to minimize downtime and improve overall system reliability.
- Conducted interviews to hire individuals based on business needs, and trained them to ensure adequate coverage and excellent customer service.
- Technologies: NetApp, Unity, Azure NetApp Files, Purity, HPE Nimble, and related management tools.

ACHIEVEMENTS

Received On the Spot award for successfully delivering a data migration project.	2021
Certified with Azure Fundamentals (AZ-900) concepts.	2021
Received Best Team Player for successfully refreshing the infrastructure.	2022
Professional boot camp in UX design from DesignBoat school.	2022