TANISHQ D. KAUSHIK

DC Metropolitan Area • (325) 203-0546 • tanishqduttkaushik@gmail.com • Portfolio • LinkedIn

EDUCATION

University of Maryland, College of Information Studies — **College Park, MD**
Master of Science (M.S.), Information Management (Data Science Specialization) (GPA: 3.8/4.0) — **May 2023**
Bachelor of Science (B.S.), Information Science (Data Science Specialization), Minor: Statistics, — **May 2021**
Dean's List for 3 semesters

TECHNICAL SKILLS

- **Data Science**: Python (NumPy, pandas, scikit-learn, Requests, BeautifulSoup), SAS, R
- **Programming Languages**: Python, Java
- **Database Management**: MySQL, Amazon RDS
- **Cloud Computing**: Microsoft Azure
- **Data Visualization**: Tableau, Power BI, AWS QuickSight

TECHNICAL EXPERIENCE

Health Data Science Projects

Machine Learning-Enhanced Predictive Modeling for Diabetes Diagnosis using R (View Full Project) — **Jan. 2023 – May 2023**

- Conducted logistic regression to predict diabetes diagnosis using age, gender, and BMI as predictors.
- Prepared the dataset by converting height and weight to SI units, calculating BMI, and recoding variables for analysis.
- Explored the dataset, revealing key statistics: 38.28% were obese, 58.06% were female, and 15.38% had diabetes, with an average age of 46.85 and BMI of 28.78.
- Visualized data with bar plots and scatterplots to understand the distribution of insurance types, smoking habits, and relationships between variables.
- Performed logistic regression, indicating that age, BMI, and gender significantly influence the likelihood of diabetes diagnosis.
- Provided coefficients and log-odds changes for each predictor.

Statistical & ML Analysis of Medical Cost Predictions using R (View Full Project) — **Jan. 2023 – May 2023**

- Conducted linear regression to predict individual medical costs billed by health insurance.
- Explored a dataset with 1338 observations across 7 variables, examining categorical and continuous variables.
- Visualized data using boxplots and scatterplots, revealing insights into the relationships between variables.
- Performed multiple linear regression, yielding coefficients for predictors such as age, BMI, gender, smoking status, number of children, and region.
- Assessed regression assumptions through diagnostic plots, revealing violations in linearity, normality of residuals, and homogeneity of residuals variance.
- Identified significant predictors, providing valuable insights for stakeholders in healthcare cost management.

Other Technical Experiences

START Database Consulting Internship — **Jan. 2023 – May 2023**

- Designed and developed a relational database using SQL for efficient storage and retrieval of information from Big Data Excel Sheet.
- Drafted an Entity Relation Diagram (ERD) schema and normalized the database to the third normal form.
- Transformed the data into tables using Python scripts and uploaded tables onto Amazon RDS using insertion scripts, allowing users such as academic researchers, government agencies, and the public to search for and access documents related to terrorism using keyword variables, thus filling the knowledge gap, and reducing redundancy in research.

Clustering Model – Credit Card Dataset (View Full Project) — **Aug. 2022 – Dec. 2022**

- Conducted a comprehensive clustering analysis on an 8600-row, 17-feature credit card dataset.
- Systematically preprocessed the data, meticulously choosing between normalization and standardization based on minimizing SSE.
- Employed the K-means clustering algorithm and tracked SSE across different cluster counts, visually illustrating the SSE vs. K relationship with an Elbow method plot and determined the optimal cluster count via the distinct elbow point.
- Visualized the clusters using a scatter plot, each cluster uniquely highlighted by color for effective data interpretation.

Classification Model - Car Dataset (View Full Project) — **Aug. 2022– Dec. 2022**

- Developed classification algorithm for car price prediction using Decision Tree and Random Forest models.
- Preprocessed car dataset by cleaning, transforming data, creating dummy variables for categorical features.
- Conducted 70-30 train-test split and evaluated Decision Tree classifier's performance using GINI and Entropy criteria.
- Created visualization of Decision Tree classifier to understand its decision-making process.
- Built Random Forest classifier using best-performing criterion and evaluated performance with accuracy scores.

TANISHQ D. KAUSHIK

DC Metropolitan Area • (325) 203-0546 • tanishqduttkaushik@gmail.com • Portfolio • LinkedIn

PROFESSIONAL EXPERIENCE

University of Maryland – Head Teaching Assistant for INST326 (OOP For Information Science) **Aug. 2022 – May 2023**

- Managed a team of 4 TAs on behalf of the professor.
- Taught Object-Oriented Programming concepts, principles, and methods.
- Taught how to design, program, test, and debug python applications.

University of Maryland – Graduate Teaching Assistant for INST414 (Data Science Techniques) **Aug. 2021 – May 2022**

- Covered the complete Analytical funnel from data extraction and cleansing to data analysis, insight interpretation and visualization.
- Taught NLP concepts and a variety of linear and non-linear classification methods.
- Graded class assignments, discussions, and hosted internal Kaggle competition among students to provide practical application of data science methods.
- Held office hours for approximately 40 students and helped with data science projects.

University of Maryland – Community Assistant **Jan. 2018 – May 2021**

- Assisted on-campus students regarding their housing and miscellaneous needs to promote an improved living experience.
- Managed the data entry and data quality of a sensitive database containing data on approximately 600 students.
- Utilized conflict resolution techniques to deescalate problems with both students and parents on campus.