

Isha Rajesh Angadi

(240) 426-9777 | ishaa@umd.edu

<https://www.linkedin.com/in/isha19/> | https://github.com/Ishaangadi/isha_resume

A summary of the major projects completed in the Masters in Data Science and Analytics at the University of Maryland, College Park is presented below.

Project Teal: Create predictions for ovarian cancer

(December 2021)

- The goal is to build a model that accurately predicts malignant or benign tumors. The Base Model was based on a research paper (*Using machine learning to predict ovarian cancer by Lu, et al.*), which analyzed data to find out if someone is at serious risk of Ovarian Cancer based on their 49 biomarkers and non-biomarkers. The team sought to improve the accuracy of the base model by tuning various hyperparameters.
- The procedure for the project consisted of handling missing values, feature selection, implementing a decision tree, and comparative analysis using logistic regression and ROMA.
- Upon utilizing stratified K-cross validation with shuffling — our team achieved 82.5% testing accuracy (with the criterias MRMR, 8 features, and entropy). When compared with the stratified K-cross validation without shuffling, our team's test accuracy with the same criteria increases by 16.7% (from 65.8% to 82.5%).
- **Technologies Used:** Python, Python Libraries, Data Analysis and Visualization

Influencer Analysis of Congress People Twitter Data

(May 2022)

- Project's objective was to provide factual information to help people decide on their political affiliations and actions by giving them a tool to match them with the views of the appropriate congressperson. Tweets from members of Congress are vast, intricate databases that are updated frequently. The velocity of the tweets was very high, approximately 26,000 tweets were sent during the course of just 10 days. Processing large amounts of unstructured, low-density Twitter data was the key challenge of the project.
- The topic modeling technique was used to better comprehend and interpret the twitter data. Python 'snsrape' was used to scrape the data from twitter, which was then sorted into a model of pre-assigned topic categories that was manually constructed by the team.

- Following the topic modeling, feature extraction with BERT was used for influencer profiling of the congressperson's tweet.
- **Technologies Used:** Python, Python Libraries, Data Science and Visualization

UCI Wine Data Exploration Application

(November 2022)

- Clean, Analyze, Visualize, and create interactive components using Wine Data set.
- Multiple tabs were created within the app
 - Tab 1: an interactive table with a search bar is present to actively look through the data
 - Tab 2: an interactive histogram with a search bar is present to actively look the attributes of the data
 - Tab 3: a correlation plot to explore the correlation between continuous variable, a linear regression model for Magnesium and confidence intervals to explore both sections present
- **Technologies Used:** R Programming, R shiny

Workout Analysis with Visual Pose Correction

(December 2022)

- The aim of the project is to create a human pose estimation and workout analyzer that assists in correcting user poses during their workouts, thus, reducing long-term injuries. The project pipeline will be inspired from *Real Time Indoor Workout Analysis Using Machine Learning & Computer Vision* (A. Nagarkoti et. al)
- The team uses an amalgamation of techniques such as 2D pose estimation, optical flow, DTW (Dynamic Time Warping), and affine transformation. The project aims to build upon these techniques to provide a more computationally efficient workout analysis.
- The project is currently in progress — a video recording of a live demonstration where the user will utilize the project to correct their exercise form.
- **Technologies Used:** Python, Python Libraries, Data Science and Visualization

Reinvent the Wheel 2.0 Hackathon Held by Torqata

(North Carolina, November 2022)

- Challenge #1: Forecasting Scrap Tires (*How many tires of a given size_code will be collected by a given Distribution Center every day?*)

A stacked LSTM (long short-term memory networks) model architecture was used on exploring different features including — cost, tire size, width, retail price, day of week, day of month, and month.

- **Technologies Used:** Python, Python Libraries, Data Science and Visualization

- Challenge #2: Non-linear Optimization (*How to efficiently move the scrap tires downstream to the recycler?*)

Using Python SciPy function minimization with SLSQP (Sequential Least Squares Programming) — the given data was converted into graph format to obtain the weights of edges to solve the optimization problem.

- **Technologies Used:** Python, Python Libraries, Data Science and Visualization