

TAMU DATATHON

Challenge Chosen: TD for you page

Team Name: Gryffindor

Members: Aashrith sai, Naga Prasanth, Vishal Indivar.

GOAL: Create a for you page that is customized for each attendee using tamu datathon application.

We have created an individual specific For-you page which is completely dynamic and constantly keeps updating as there are more entries into the database. We have created this application keeping these key features in mind.

Key features:

1. Team members recommendation
2. Workshops recommendation
3. chatbot that answers all the queries
4. Real time graph visualizations

Implementation specifics:

Team members recommendation:

An ideal team has team members from similar backgrounds who have varied skills required to complete a project. So for our website we developed a team member recommendation system by first finding the similar individuals. To find similar individuals we used school name (individuals from same school stay at close proximity and may know each other), degree pursuing, majors (People from similar backgrounds would like to work on similar problems), data science experience, number of hackathons attended (as projects are classified easy, medium, hard people with similar experience choose same projects) as features. We first encoded these categorical attributes using one-hot encoding and then clustered these individuals based on the above mentioned features to find highly similar individuals using cosine similarity. We then further improved our system by matching individuals with high similarity and having complementary skills as people with different skills are required to complete a good project. We tried to recommend people to maximise the skill-set of the team as a whole.

0c554e8cf1ddb8fbd062820829048cf0

Education:

School: Texas A&M University

Classification: Masters

Major: Management Information Systems

Experience:

Datascience Experience (yrs): 2

Hackathons Attended: 0

Fig-1: Details of the Individual suggested.

For creating the Team member recommendation system we used sklearn to do one-hot encoding, finding cosine similarity and matplotlib to visualize the data and find correlations. This intelligent team building application takes care of matching individuals who can complement each other with their skills. It is often the case where someone with good skills with machine learning needs to be complemented by a web developer. Keeping this in mind we have created this recommendation engine.

Your Dream Team

Below are our Top 5 Recommendations for you, to consider as your potential teammates, go ahead connect with them on discord and other platforms!

0c554e8cf1ddb8fbd062820829048cf0	+
3f1c76ad8b6651fdfabbc3322aee0b04	+
ee257d31fc9469d2ca2c6ca629df35f0	+
802fb86c93296ae610ba81e48b5b5785	+
9b435d7fdce819358ab0dd7166032677	+

Fig-2: Recommendation of top 5 team mates to make a complete team.

Workshops recommendation:

We also created a personalised workshop recommendation system that recommends workshops based on the students' expertise and difficulty of workshops using the “workshops”, “Applicant Data” datasets given to us. For this we first clustered the individuals as beginner, intermediate and advanced by using their data science experience, the number of hackathons attended provided to us when the user signs in. We then recommended the workshops based on the difficulty of the workshop and also their interests. This recommendation system helps individuals find best workshops that are suited to their skill set and also have the opportunity to learn. We have also displayed relevant information about the workshops, skills targeted, venue and time of the workshop. This would save a lot of time.

Introduction to Python	-
<p>Are you new to programming? Or do you want a refresher to brush up on what you know? No worries! In this class, we will learn the fundamentals of python. From variables, to data types, to loops, we will cover the basics of the most widely used programming language in data science!</p> <p>Track: CS</p>	
Storytelling with Data	+
Stats for Data Scientists	+
How To Win TAMU Datathon	+

Fig-3: Workshop recommendations based on interest and difficulty

Reveille the Chatbot:

We created Reveille to make the website more lively and give users an interactive experience. We implemented a chatbot that answers user queries regarding workshop suggestions, hackathon sponsors, prizes, competitions etc. For implementing the chatbot we used the workshops dataset to get the relevant information about each workshop and then the tamu datathon website to get information regarding the hackathon. The chatbot needs keywords and phrases to respond as it is rule-based, we referred and scraped websites to get relevant keywords for industries, technologies etc. For implementing the chatbot we used TensorFlow deep learning framework to create and train the neural network, NLTK library to text preprocessing (stemming, lemmatization, removing stop words, removing punctuation), and a custom bag of words function in python. We then used the bag of words approach to convert text to features and trained a feed forward neural network (FNN) that determines the class of question. Based on the output of the neural network the chatbot program determines the subset of answers from which it returns a relevant answer as output. For testing the chatbot experience we used the queries dataset to determine if the workshop details for the questions present were answered perfectly without any error. We then also tested and trained it with several queries to make it robust and relevant. Reveille makes the user experience whole and complete, it answers questions from the date and time of the workshops to suggesting challenges, talks and workshops to work on one's skills. It is a one stop shop for anyone who is participating in the datathon.

Ask Me Anything!

Hi, I am Reveille, I am here to help you with any queries you may have regarding workshops, Sponsors and much more!

you:

I am interested in Natural Language Processing

Interested in learning what Natural Language Processing is all about? The hour long Natural Language Processing course from 17-18.15 CST will teach you how to identify and separate words, extract topics, and conduct sentiment analysis. See you there!!

Fig-4: Reveille lets users learn about workshops of their interest and also makes meaningful conversations

Ask Me Anything!

Hi, I am Reveille, I am here to help you with any queries you may have regarding workshops, Sponsors and much more!

you:

hello!

Good to see you again!

Fig-5: Conversations with reveille.

Real Time Graph Visualizations:

We have also added several real time graph analysis tools, where one can filter on what criteria the data should be visualized. We included categories such as students ratio from different colleges, classification, expertise and industry from which they are from. This analysis tool helps individuals to have a better understanding of their peer group and can be well prepared about their competition. We have used a machine learning model to categorize individuals into beginner, intermediate and advanced based on their inputs, every time new records are added to the database, real time visualization takes place. There is no time latency or there is no requirement of additional computations to create visualizations.

Pie Chart

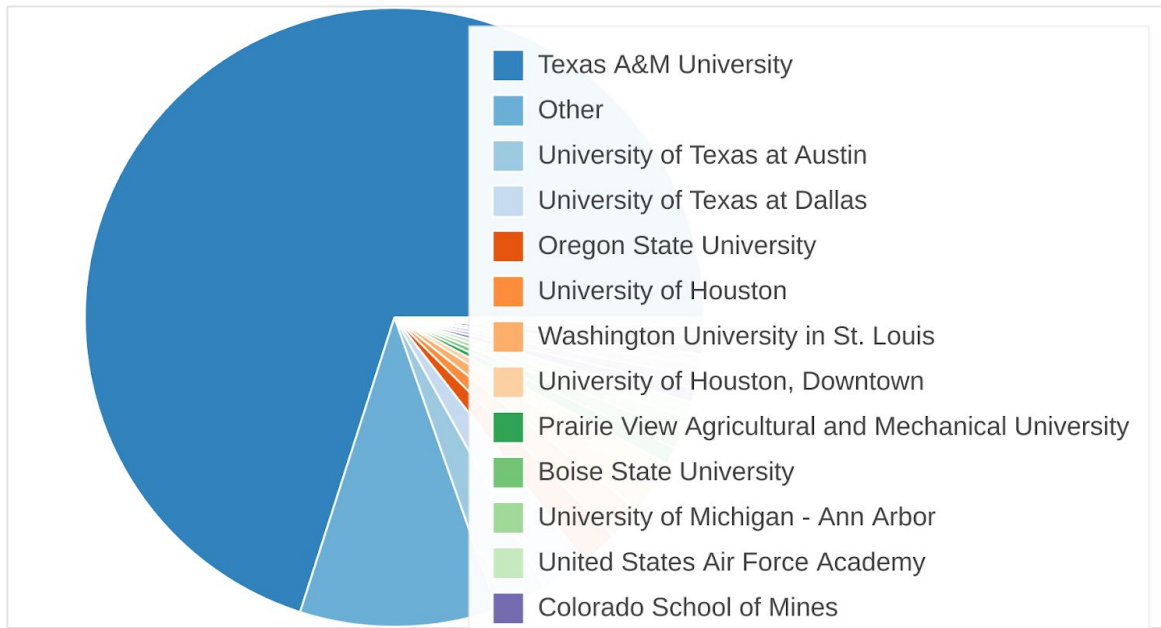


Fig-6: Visualization based on college.