KHANJAN PUROHIT

Bangalore, Karnataka, India-560004 | +91(910)623-6419 | khnjn.purohit@gmail.com
linkedin.com/in/khanjanpurohit | github.com/khanjan18

Hard-working Data Science and Analytics major (8.6 CGPA) currently attending Jain (Deemed-to-be University). Aiming to leverage proven analytical, creative thinking, and problem-solving skills to successfully fill the data scientist role at the firm.

EDUCATION

Jain (Deemed-to-be University), Bangalore, India   *June 2019 - May 2022 (Expected)*
Bachelors Of Science (Honours.) - Data Science and Analytics   **CGPA - 8.6**
Courses : Machine Learning, Deep Learning, Python, Natural Language Processing, Data Preprocessing, Data Mining EDA, SQL, Probability Distribution, Linear Algebra, Database Management System, Business Analytics.

PROJECTS

Movie Recommendation System   *January 2022 - Current*
- Built a recommendation system website mrs-kp.herokuapp.com which recommends the movies like the input movie with the details of the input movie (poster, genres, casts, runtime etc.). There are a total of 6,308 hollywood movies from 1916 to 2021 in the system.
- First, data is collected from Kaggle and Wikipedia using web scraping. Data is preprocessed. Cosine Similarity is used for the creation of the recommendation system. Front-end and Back-end work of the website is being handled using the Streamlit library of Python, CSS, HTML. At last the model is deployed using Heroku Deployment.
- After searching for the movies on the website, there will be the details of the movie and cast details such as profile, name, character name etc. It fetches these details from TMDB API. There will be 8 movies recommended with name and poster.

Stock Sentiment Analysis Using News Headlines   *Nov 2021 - Dec 2021*
- Analyzed the dataset and created ML RandomForestClassifier model with 86% accuracy after preprocessing the data.
- In preprocessing, data are split into training and testing sets according to date because it is a time series dataset. Headlines are processed through TfidfVectorizer.

Natural Language Processing With Disaster Tweets   *Sept 2021 - Oct 2021*
- Participated in a kaggle competition "Natural Language Processing With Disaster Tweets", analyzed data and classified whether the tweet is a disaster or not.

PUBLICATION

Separation Of Data Cleansing Concept From EDA   *May 2021 - June 2021*
- Published an article "Separation Of Data Cleansing Concept From EDA" in International Journal Of Data Science And Analysis, Vol.7, No.3, doi: 10.11648/j.ijdsa.20210703.16