

Kausik Chattapadhyay

Data Science Projects

Fraud Detection for Transactions

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Fraud%20Detection%20For%20Transactions>

Development of a machine learning model to predict whether a financial transaction is fraudulent or legitimate. The Fraud Company aims to enhance its fraud detection service in the Brazilian market. With a revenue-sharing model heavily dependent on correct fraud detection, the company faces significant financial risk if its models are not precise. This project aims to mitigate this risk by building a robust machine learning model capable of accurately finding fraudulent transactions.

Credit Card Approval

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Credit%20Card%20Approval>

The goal of this project is to define a fraud mentoring model to help banks to automate the Credit Card eligibility process based on customer detail provided while filling online application form & Credit history of customer. The decision of approving a credit card or loan is majorly dependent on the personal and financial background of the applicant. Factors like, age, gender, income, employment status, credit history and other attributes all carry weight in the approval decision. Credit analysis involves the measure to investigate the probability of a third-party to pay back the loan to the bank on time and predict its default characteristic. Analysis focus on recognizing, assessing, and reducing the financial or other risks that could lead to loss involved in the transaction.

Daily Slack updates integrating Flight and Weather APIs with data sources

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Daily%20Slack%20updates%20integrating%20Flight%20and%20Weather%20APIs%20with%20data%20sources>

In modern application development, integrating data from diverse sources in various formats is a common challenge, especially in the realm of machine learning applications. Apache NiFi presents a powerful solution for seamlessly orchestrating data flow between different systems. This project demonstrates how Apache NiFi can be leveraged to collect data from two distinct APIs - aviationstack.com for real-time flight data and visualcrossing.com for weather data - and subsequently write this data to multiple databases while also sending daily notifications to Slack. The project provides a step-by-step research guide on setting up data ingestion, transformation, and dissemination pipelines using Apache NiFi processors.

Artist identification from their Paintings.

<https://github.com/chatkausik/Second-Capstone-Project>

Artist identification is traditionally performed by art historians and curators who have expertise and familiarity with different artists and styles of art. This is a complex and interesting problem for computers because identifying an artist does not just require object or face detection; artists can paint a wide variety of objects and scenes. This approach is motivated by the hypothesis that every artist has their own unique style of art and that we can improve upon existing artist identification methods by using a CNN to determine the best possible feature representation of paintings.

IMDB Movie score recommender & Movie review sentiment analysis.

<https://github.com/chatkausik/First-Capstone-Project>

1. What can we say about the success of a movie before it is released? Are there certain companies (Pixar?) that have found a consistent formula? Given that major films costing over \$100 million to produce can still flop, this question is more important than ever to the industry.

2. This question puzzled almost everybody for a long time since there is no universal way to claim the success of movies. Many people rely on critics to gauge the quality of a film, while others use their instincts. But it takes the time to obtain a reasonable amount of critic's review after a movie is released. And human instinct sometimes is unreliable.

3. Predicting IMDB Score of a movie before it released in cinemas is my primary goal for this project without relying on critic's review data and human instincts data along with movie review sentiment analysis.

4. This will benefit all cinema lovers like me or film producers/directors who can get a high of overview of the probable score of the new movie they are releasing like a pre-poll forecast.

* How to recommend IMDB Movie Score(Score ranges from 0-Low to 10-Excellent) for the new movies?

* How to identify bad (0) or good(1) sentiment from Movie reviews?

Retail Inventory Management

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Retail%20Inventory%20Management>

Goal: Building a binary classifier which gives us a list of product ID which need to be retained in the inventory or list of products that need to be removed.

Context: A retail firm has many products in their inventory, and very few of them tend to sell (only about 10% sell each year) and many of the products only have a single sale in the course of a year.

Objective: The sales and growth team of the retail firm wants to determine which products from their inventory should they retain to sell and the ones to discard.

Data: The data given contains both historical sales data AND active inventory.

Movie Recommendation System

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Movie%20Recommender%20System>

The goal of this project is to provide a recommendation system for video content providers to predict whether someone will enjoy a movie based on how much they liked or disliked other movies.

Recommendation Systems are a type of information filtering systems as they improve the quality of search results and provides items that are more relevant to the search item or are related to the search history of the user.

They are used to predict the rating or preference that a user would give to an item. Almost every major tech company has applied them in some form or the other. Major companies like YouTube, Amazon, Netflix use recommendation systems in social and e-commerce sites use recommendation system for its users to suggest for an individual according to their requirement more precise and accurate. These online content and service providers have a huge amount of content so the problem which arises is which data is required for whom so the problem of providing apposite content frequently. This project represents the overview and approaches of techniques generated in a recommendation system.

MLB Attendance

<https://github.com/chatkausik/chatkausik.github.io/tree/master/MLB%20Attendance>

* Use MLB data to make a recommendation on how to improve attendance

* Data was obtained from the Los Angeles Dodgers Major League Baseball team

Covid 19 Data Analysis and Comparison

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Covid%2019%20Data%20Analysis%20and%20Comparison>

The novel coronavirus, also known as SARS-CoV-2, is a contagious respiratory virus that first reported in Wuhan, China. On 2/11/2020, the World Health Organization designated the name COVID-19 for the disease caused by the novel coronavirus. This new strain of virus has strike fear in many countries as cities are quarantined and hospitals are overcrowded. This project aims at exploring COVID-19 through data analysis and visualization.

House Price Prediction

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Housing%20Price%20Prediction>

The objective of this project is to build a Machine Learning model for the prediction of housing prices based on pattern extracted from our USA Housing dataset which contains data about USA housing sales provided by Kaggle. This dataset contains features like number of bedrooms, number of bathrooms, sqft, year built, year renovated, condition, zip code and the target variable will be price. We will use various visualization methods to analyze the correlation of each feature as part of EDA. Descriptive analysis and quantitative analysis will use subsets of it depending on models. This project consists of two phases: -

- * Phase I: Focuses on data cleaning, exploration and preprocessing.
- * Phase II: Machine Learning model building, validation and prediction.

Airline Safety

<https://github.com/chatkausik/chatkausik.github.io/tree/master/Airline%20Safety>

Is the Airline safe?

Aviation industry is highly sophisticated, very competitive and constantly evolving. We will be amazed if we think about the emerging and evolving innovations regarding comfort and safety in the airlines space. Ranging from evolving cockpit technology, constant flight status monitoring, passenger cabin safety measures, ultra-modern communication through air traffic control and airport control, security checking and airport safety, - all work together to provide a better and safer travel experience. Considering all the aspects, air travel is arguably the safest means of travelling in the present era.

