Capstone Projects Overview

Project 1: Time Series-Carbon Emissions Forecasting

Context

Global warming is one of the foremost problems that humanity needs to solve to ensure our survival for the future. There is enough scientific evidence to suggest that there could be catastrophic effects if we don't start controlling the amount of damage we do to our environment and ecosystem.

One of the primary contributors to global warming is the accumulation of greenhouse gases in our atmosphere. These greenhouse gases are generated as a result of various industrial and economic activities that we conduct in our day-to-day lives and CO_2 is one of the biggest contributors to this greenhouse effect. The need of the hour is to be able to identify the sources of CO_2 and see how to assess the impact and potential ways to control the emission.

A lot of electricity in the US is still generated by various sources of energy that are burned to run turbines to generate electricity, and the burning of these fuels is the primary source of CO_2 emissions. Different approaches and technological advancements are being used to minimize the emission of carbon & trade-off between carbon emission and affordable electricity production for mankind. Forecasting CO_2 emissions can make an impact on decision-making in terms of emission reduction & choosing better methods of electricity production.

Objective

Forecast the carbon emissions value for **natural gas (NNEIEUS)** fuel type for the next 12 months and propose certain measures that can be adopted as policies to reduce these emissions.

Project 2: Regression- Used Cars Price Prediction

Context

There is a huge demand for used cars in the Indian Market today. As sales of new cars have slowed down in the recent past, the pre-owned car market has continued to grow over the past few years and is now larger than the new car market. Cars4U is a budding tech start-up that aims to find footholes in this market.

In 2018-19, while new car sales were recorded at 3.6 million units, around 4 million second-hand cars were bought and sold. There is a slowdown in new car sales and that could mean that the demand is shifting towards the pre-owned market. In fact, some car owners replace their old vehicles with pre-owned cars instead of buying a new automobile.

Unlike new cars, where price and supply are fairly deterministic and managed by OEMs (Original Equipment Manufacturer / except for dealership level discounts which come into play only in the last stage of the customer journey), the used car market is a very different beast, with large uncertainties in both pricing and supply. Several factors, including mileage, brand, model, year, etc. can influence the actual worth of a car. From the perspective of a seller, it is not an easy task to set the correct price of a used car. Keeping this in mind, the pricing scheme of these used cars becomes important in order to grow in the market.

Objective

Come up with a pricing model that can effectively predict the price of used cars and can help the business in devising profitable strategies using differential pricing.

Project 3: Recommendation Systems-Music Recommendation System

Context:

With the advent of technology, societies have become more efficient with their lives. At the same time, however, individual human lives have also become more fast-paced and distracted, leaving little time to explore artistic pursuits. Also, technology has made significant advancements in the ability to coexist with art and general entertainment. It has in fact made it easier for humans with a shortage of time to find and consume good content.

Almost every internet-based company's revenue relies on the time consumers spend on its platform. These companies need to be able to figure out what kind of content is needed in order to increase customer time spent and make their experience better. Therefore, one of the key challenges for these companies is figuring out what kind of content their customers are most likely to consume.

Spotify is one such audio content provider with a huge market base across the world. With the ever-increasing volume of songs becoming available on the Internet, searching for songs of interest has become a tedious task in itself. However, Spotify has grown significantly in the market because of its ability to recommend the 'best' next song to each and every customer based on a huge preference database gathered over time - millions of customers and billions of songs. This is done by using smart recommendation systems that can recommend songs based on users' likes/dislikes.

Objective

Build a recommendation system to propose the top 10 songs for a user based on the likelihood of listening to those songs.

Project 4: CNN- Malaria Detection

Context

Malaria is a contagious disease caused by **Plasmodium parasites** that are transmitted to humans through the bites of infected female Anopheles mosquitoes. The parasites enter the blood and begin damaging red blood cells (RBCs) that carry oxygen, which can result in respiratory distress and other complications. The lethal parasites can stay alive for more than a year in a person's body without showing any symptoms. Therefore, late treatment can cause complications and could even be fatal. Almost 50% of the world's population is in danger from malaria. There were more than 229 million malaria cases and 400,000 malaria-related deaths reported over the world in 2019. Children under 5 years of age are the most vulnerable population group affected by malaria; in 2019 they accounted for 67% of all malaria deaths worldwide.

Traditional diagnosis of malaria in the laboratory requires careful inspection by an experienced professional to discriminate between healthy and infected red blood cells. It is a tedious, time-consuming process, and the diagnostic accuracy (which heavily depends on human expertise) can be adversely impacted by inter-observer variability.

An automated system can help with the early and accurate detection of malaria. Applications of automated classification techniques using Machine Learning (ML) and Artificial Intelligence (AI) have consistently shown higher accuracy than manual classification. It would therefore be highly beneficial to propose a method that performs malaria detection using Deep Learning Algorithms.

Objective

Build an efficient computer vision model to detect malaria. The model should identify whether the image of a red blood cell is that of one infected with malaria or not, and classify the same as parasitized or uninfected, respectively.

Project 5: Classification-Loan Default Prediction

Context

A major proportion of retail bank profit comes from interests in the form of home loans. These loans are borrowed by regular income/high-earning customers. Banks are most fearful of defaulters, as bad loans (NPA) usually eat up a major chunk of their profits. Therefore, it is important for banks to be judicious while approving loans for their customer base.

The approval process for the loans is multifaceted. Through this process, the bank tries to check the creditworthiness of the applicant on the basis of a manual study of various aspects of the application. The entire process is not only effort-intensive but also prone to wrong judgment/approval owing to human error and biases.

There have been attempts by many banks to automate this process by using heuristics. But with the advent of data science and machine learning, the focus has shifted to building machines that can learn this approval process and make it free of biases and more efficient. At the same time, one important thing to keep in mind is to make sure that the machine does not learn the biases that previously crept in because of the human approval process.

Objective

Build a classification model to predict clients who are likely to default on their loan and give recommendations to the bank on the important features to consider while approving a loan.