National College of Ireland

Enhancing Irish Tourism : A Detailed Approach Integrating Weather Prediction Models and Recommendation Systems

MSc Research Project
Data Analytics

Mohamed Mubassir Hussain Hidayat Hussain
Student ID: X21227471

School of Computing
National College of Ireland

Supervisor: Abdul Shahid

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Mohamed Mubassir Hussain Hidayat Hussain
Student ID:	X21227471
Programme:	Data Analytics
Year:	2023
Module:	MSc Research Project
Supervisor:	Abdul Shahid
Submission Due Date:	31/01/2024
Project Title:	Enhancing Irish Tourism : A Detailed Approach Integrating Weather Prediction Models and Recommendation Systems
Word Count:	8450
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Mohamed Mubassir Hussain Hidayat Hussain
Date:	31st January 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Irish Tourism : A Detailed Approach Integrating Weather Prediction Models and Recommendation Systems

Mohamed Mubassir Hussain Hidayat Hussain
X21227471

Abstract

Tourism Industry is one of the vastly growing sector in Ireland and is one of the biggest contributors to the economy of Ireland. The tourism Industry in Ireland is heavily impacted by changes in the seasons, which also has a major effect on the number of visitors and the overall economic state. With the help of advanced algorithms and the usage of a historical data and a detailed weather analysis to develop a recommendation systems for tourists to give them customised suggestions on the places to visit in different times of the year based on historical data and current trends ensuring a dynamic and responsive system while also developing an weather prediction model with for accurate weather forecasts. In this research, using the content based recommendation systems it successfully provides customised suggestions to tourists based on the historical data and the weather conditions. Along with that the Random Forest regression model used for weather prediction shows great accuracy in predicting the weather. Overall this research gives great results with the random forest regression giving out high accuracy in weather prediction and the content based recommendation provides personalised suggestions to the tourists based on the historical data. These results provide valuable information to the tourists as well as the industry stakeholders to enhance the Irish tourism sector.

1 Introduction

Tourism in Ireland is one of the major contributors to the economy of Ireland with millions of people visiting Ireland every year generating a huge sum of revenues as well as employment opportunities. Yet, the tourism industry struggles with the challenges related to seasonality of tourist activities, impacting local businesses and infrastructure. The paper by Li et al. (2019) analyses in the field of text corpus-based tourism big data mining. It discusses the techniques and difficulties involved in getting useful informations in the tourism industry. This research establishes a strong base for understanding the importance of analysis of language for improving recommendation systems. This research is very important to our research project, providing important suggestions to create a smart recommendation systems. Then the paper by Wang et al. (2016) analyses the integration of artificial intelligence in order to improve the accuracy of models used to estimate residential space heating.The innovative approach presented in this research aligns with our project's focus on weather prediction models and can potentially offer insights into developing an accurate weather prediction model for the tourism industry.

1.1 Significance, Motivation and Objectives of this Research

Understanding knowledge about the impact of seasons and climate on tourist behaviour is very important for the development of efficient approaches to enhance the overall tourism experience. The Ireland tourism industry experiences a similar issue i.e, it sees an increase in the number of tourists throughout the summer, making the attractions to reach their maximum capacity. While, the extreme cold weather during winters reduces the number of many tourists. The purpose of this research is to support Ireland's tourism sector against economic fluctuations caused by unpredictable variations in the weather and season and the strong motivation behind this research is not just to address the seasonality in Ireland but also to make sure that with the help of this research there is economic stability throughout the country especially in the tourism industry and also this research will promote off season travel in order to prevent the temporary closures of businesses and job cuts. This research has two primary objectives. The research aims to analyse patterns and trends that impact tourist decisions through the integration of various data-sets that include information on accommodation, activities, attractions, and weather. The integration of machine learning techniques, such as content-based filtering to develop an adaptable and interactive recommendation system for tourists as well as usage of Regression techniques to predict the weather accurately. The approach attempts to provide customised recommendations for throughout the year places to visit through analysing historical data and predict weather conditions. After reviewing existing literature works related to my research like Abbasi-Moud et al. (2021) and Arif et al. (2022) as well as assessing their gaps in current methods to deal with seasonality, weather and a personalised tourist recommendation system in the tourism sector it was very evident there was a need for innovative solutions to address these issues.

1.2 Research Question

How can the usage of machine learning and weather prediction models improve tourist recommendations, resource allocation, and travel patterns in the terms of Irish tourism ?

1.3 Structure Of The Report

The Report has the following structure : Section 2, we discuss a detailed literature review of research papers. Section 3, we discuss about the research process flow and its stages. Section 4 we discuss the implementation the project. Section 5 we discuss the evaluation part of the research. Section 6 we discuss how the model performs. Section 7 we discuss the conclusion and Future works of the research.

2 Literature Review

In this section, we perform an in-depth analysis of a set of research papers which are relevant to our research project. Through a thorough examination of existing research papers, our goal is to place our research within the big scope of tourism, machine learning applications, and weather prediction models. This study not only addresses issues and limitations in existing research but also provides a foundation for our approach. By analysing the existing work, we seek to gather important information, thoughts, and

observations that will improve our approach and improve our understanding in this area of study. Below it shows the list of papers referred for this research in Table 1

Table 1: Research Papers

Paper Title	**Year**	**Strength**	**Weakness**
Cafob: Context-aware fuzzy-ontology-based tourism recommendation system	2022	Uses advanced fuzzy ontology	Limited scalability
Tourism recommendation system based on semantic clustering and sentiment analysis	2021	Usage of sentiment analysis for recommendations	Semantic clustering can cause oversimplify
Destinations ratings based multi-criteria recommender system for Indonesian halal tourism game	2022	Consideration of multiple scenarios for recommendations	Limited to Only Indonesia
Using poi functionality and accessibility levels for delivering personalized tourism recommendations	2019	POI functionality for personalised recommendations	Can avoid less popular locations
AI-based campus energy use prediction for assessing the effects of climate change	2020	Usage of AI for energy use prediction	Relies on specific data
How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management	2021	A detailed review of smart tourism recommendation systems	Less efficiency
Deep learning-based tourism recommendation system using social network analysis	2020	Usage of social network for tourism	Limited usage of deep learning models
An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data	2020	Usage of AI techniques	Can cause data limitations
A review of text corpus-based tourism big data mining	2019	Gives insights in text corpus-based tourism big data mining	Less efficiency
Leveraging geospatial data in python with geopandas	2023	Explains practical applications well	Limited scalability and performance
Applying internet information technology combined with deep learning to tourism collaborative recommendation system	2020	Combines internet tech and deep learning for collaborative recommendations	Handling of large scale datasets
Artificial intelligent models for improved prediction of residential space heating	2016	Usage of AI techniques	Lack of clarity on models and user acceptance
Navigating through the complex transport system: A heuristic approach for city tourism recommendation	2020	Unique approach for tourism	Not effective enough to handle city tourism

Continued on next page

Table 1 – Continued

Paper Title	Year	Strength	Weakness
Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction	2021	Impact of AI on weather forecasting and climate monitoring	Limitations in the usage of AI in weather prediction

2.1 Dataset Exploration

Dataset Exploration is one of the key aspects of my research project given that there were four data-sets which were of different sizes and different data present. It was a challenge to address that but the research paper by Li et al. (2019) had a heavy impact on my project by giving a detailed understanding of approaches and shortcomings present in processing various types of data within the tourism industry. This research's work has laid a foundation for the better understanding of the data-sets and data preprocessing techniques that are very much relevant to data's within the tourism industry. The insights helped a lot in understanding the complexities in managing multiple data-sets and then Rafaat (2023) also played a key role in the Data Integration part of our research project by providing deep insights into merging our different data-sets using various attributes and spatial joins with which it performs a nearest-neighbor search, merges the data-sets based on the nearest points.

2.2 Tourism Recommendation Systems:

Recommendation systems, commonly known as recommenders, are a bit computational algorithms very much designed to suggest places or content to users based on their preferences, behaviors, or similarities with other users. Here dos Santos et al. (2019) proposes a recommendation system based on users physical or psychological limits and also offers models and algorithms for a tourism recommendation system based on user and POI characteristics. By using Points of Interest (POI) functionality and accessibility levels, It excels in personalised tourism recommendations, giving a brief understanding of personalised suggestions to individual preferences. Jeong et al. (2020) examines internet tourism data, including visitor and user feedback. It deeply analyses tourist terms using social network analysis to identify frequency and associations. The main purpose is to obtain tourist attraction data from huge data-sets and also uses tourist attraction criteria and relationships to highlight user-friendly recommendations and finally it gives a clear idea about tourism recommendation system by showcasing ways to use analysis of social networks as well as deep learning for more customised and user-friendly recommendations.Zheng et al. (2020) addresses city tourism transport mode selection difficulties to improve the smart tourism recommendation system.It presents an approach to develop personalised daily plans which include motion methods with an improved particle swarm optimisation and differentiation algorithm and other sorting algorithms were developed to achieve this. To assess their approach, they conducted a Chengdu case study. They found that their approach produces more sensible, diversified, and customised itineraries than others. This approach suggests that it could improve digital strategy and tourist experiences. Wang (2020) examines the ever-changing nature of personalised travel in the tourist industry and recognises the drawbacks of old service models. It proposes a deep learning methodology to classify tourism product information to improve service

and personalisation. Word integration in data preprocessing, Convolutional Neural Network (CNN) for processing user reviews and tourism service items, Deep Neural Network (DNN) for handling information, and factorization machine technology to improve the prediction model. It improves the field of tourism service industry model by providing insights to create a more personalised tourism service system that meets the tourists preferences.Abbasi-Moud et al. (2021) focuses on recommendation systems and fixes a limitation in which many excisting mobile tourism recommender systems fail to use information, evaluations, and ratings from other visitors with similar interests.iTravel a proposed solution by this research focuses on improving attraction suggestions using mobile communication by the tourists. Three data communication methods are used, and certain experiments are conducted to identify their performances. Also, a user study is undertaken to assess the usage of the system.Hamid et al. (2021) The e-tourism management and recommender systems in smart tourism are deeply analysed. This particular research analyses 65 papers on smart-based tourist recommender systems (TRS) and tourism marketing from the past eight years of literature. Collaborative filtering, content model, context model, and hybrid model smart-based TRS all the models have been briefly analysed. Out of all the models The content model-based approach has a great impact on smart e-tourism (40.2 percentage mean) and also it offers new opportunities, addressing various issues, and generating new ideas for both researchers and practitioners.Arif et al. (2022) develops an innovative algorithm which uses a game to recommend cool Islamic travel spots in Indonesia. It is more like a similar friend who's helping you out . Not like ordinary recommendations, which just focuses only on a single objective, this approach considers eight factors in order to provide more detailed support. The usage of Islamic travel spots in Batu City was done and while assessing it. They got some really good results which indicated that the performance was great with an average accuracy of 0.60.Abbasi-Moud et al. (2022) In order to assist travellers in making decisions against the huge information accessible, the proposal suggests a unique Context-Aware Fuzzy-Ontology-Based Tourism Recommendation System (CAFOB). CAFOB refines user preferences by introducing sentiment/emotion scores and fuzzy-weighted ontology. Using both hybrid and language related similarity, it analyses nearby attractions based on previous assessments and offers customised suggestions. This particular approach helps improve the recommendation accuracy.

2.3 Climate and Weather Prediction

The usage of various Machine Learning models for the weather prediction takes it to a next level giving at-most accuracy and performance. Wang et al. (2016) examines the improvement of hourly domestic space heating power demand prediction through the use of artificial intelligence (AI) models an affordable electricity metre, and readily available meteorological data. The research uses four artificial intelligence (AI) models to forecast the hourly use of electricity for domestic heating support vector regression (SVR) radial basis function neural network (RBFNN) general regression neural network (GRNN) and backpropagation neural network (BPNN). The results show that SVR has done it better in predicting hourly household heating electricity demand than the other AI models. The study also explores the influence of human behaviour that is dynamic on prediction accuracy and finds that it has an decreasing effect on the AI models performance. Kim et al. (2020) explains the creation of an very accurate maize yield forecast model for the Midwest region of the United States in spite of the severe weather conditions. Satellite

pictures and weather information were used to test six artificial intelligence algorithms. By using hyper parameter optimisation, it clearly states that the deep neural network (DNN) model has performed better than previous models by 51–98 percentage during droughts and 30–77 percentage during heatwaves, as shown by root mean square error. Long-lasting heatwaves did not affect the DNN model's stability. This research provides a lot of insights for the weather prediction model by showing the major use of optimised AI models to predict agricultural production in extreme weather. Fathi et al. (2020) This study focuses on the importance of buildings in terms of energy use, particularly in industrialised nations, and the predictive power of artificial intelligence (AI). The research goes beyond to provide an energy forecast tool for campuses that takes climate change impacts and considering building attributes into account. This AI-powered programme forecasts campus building energy use with 90 accuracy, offering important information about the consequences of long-term climate change. The study, which is based on research from the University of Florida, is very much important to our weather forecast model because it highlights the useful applications of AI addressing challenges related to climate.Dewitte et al. (2021) showcases how artificial intelligence (AI) is being a great tool for forecasting, weather prediction, and climate monitoring. Weather forecasting efficiency depends on AI's capacity to examine massive, unstructured data-sets without the help of any specific complex processes. Proper usage of computer resources and less reliance of human labour and mean while improving the forecasting accuracy are all done perfectly by AI . The major understanding that AI plays a key role in the prediction of weather and climate analyses in the coming future. This helps in understanding and improving the efficiency and accuracy of the weather forecast model.

2.4 Conclusion on Research Papers

This detailed literature survey on both the recommendation systems and weather predictions gives us a clear insights of about the development of tourism recommendation systems and the applications in weather prediction also addressing various methodologies, challenges and innovations in various domains. Each research paper has contributed to the research like Li et al. (2019) work on text corpus based data mining helped us to understand the various methodologies to enhance the recommendation systems based on historical data. Then with reference from Hamid et al. (2021) It clearly gives this research a major insight about the usage of recommendation systems and especially the author highlights the use of content based recommendation systems for a better recommendations for textual data.With this detailed research study on various research papers it clearly states that there are still some gaps to be filled especially when it comes to enhancing tourism industry with the means of integration weather prediction with recommendation systems. Therefore it makes more sense to develop a suitable recommendation system along with a weather prediction model to enhance the Irish tourism Industry with the knowledge obtained from the literature survey.

3 Methodology

[1] This research project makes use of certain machine learning techniques to develop recommend-er systems for the user based on the data that is obtained and to predict

[1]https://www.javatpoint.com/kdd-vs-data-mining

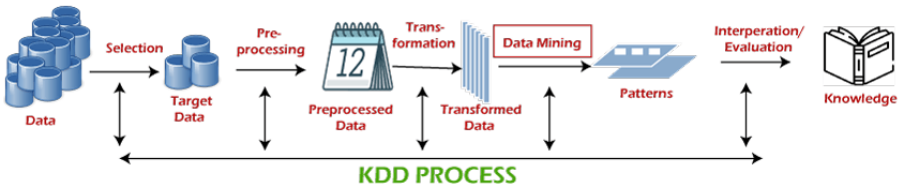

Figure 1: KDD Approach Flow Chart

the weather in various parts of Ireland. The major goal of this research project is to make sure that users have a better understanding of the places to visit in Ireland as well as the weather at different times of the year, which will enhance the overall experience of tourists.Li et al. (2019) helps us in understanding the research better and with that information we are able to go ahead with the KDD approach for our research project. The Fig. 1 shows that we have adapted to the Knowledge Discovery in Databases (KDD) method in our research project. This technique provides a method that is both organised and comprehensive, beginning with the collecting of data and moving along through the construction of models and evaluation.
In addition to addressing the process of acquiring data from Kaggle for Weather and the Failte Ireland Open Data platform for activities, attractions and accommodations, the methodological approach also addresses various elements like the additional data preparation and feature engineering. Finally, a number of different machine learning techniques will be studied, and meteorological data will be merged in order to address issues over seasonality and to give a user recommendation system that encourages travel during off-season.

3.1 Overview

The Fig. 2 clearly explains the road-map of the research project which gets started from acquiring Raw Data which has all the data-sets utilized in this project which were obtained from different sources, giving us a much detailed perspective on Irish tourism dynamics. Activities, attractions, accommodations data-sets were obtained from Failte Ireland Open Data Platform whereas the Weather data-set was obtained from Kaggle making sure that the data's are of varied and relevant information.Then it is followed by the careful progress of Data preprocessing for all the data-sets which is a bit dyanamic process given that we have 4 data-sets in the research project to deal with. Then the further steps in the research would be Data Integration which is a key aspect of our research and is the one which joins all the data-sets as one using some common factors and Feature engineering which is helpful in creating relevant features to further analyse. These are are the key elements of the research and Finally concludes in the implementation and assessment of efficient recommendation systems as well as weather prediction models. The Process flow clearly states the effort taken to use the potential of certain machine learning models in order to address the complexities of Irish tourism, particularly with regard to addressing the issues that are brought by seasonality through our approach using recommendation systems and weather prediction models.

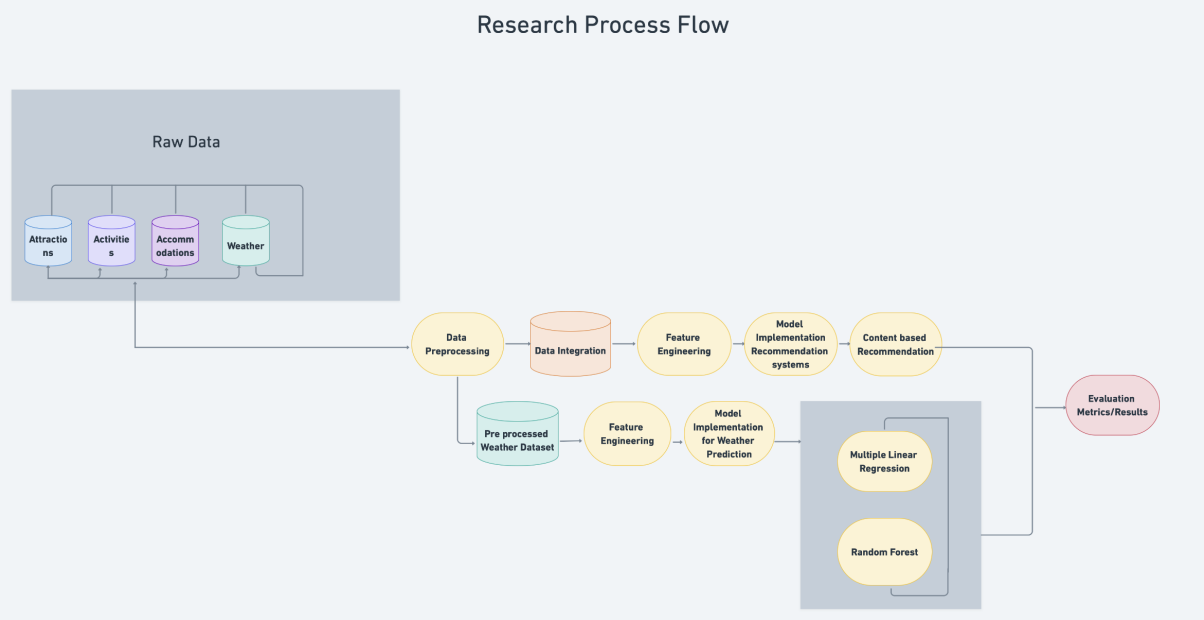

Figure 2: Project Design Flow Chart

3.2 Data Description

Column	Description
Name	Name of the place/attraction/activity
URL	Website link of the place/attraction/activity
AddressRegion	County In Ireland
AddressLocality	Denotes the Country
Tags	Denotes the type of destination
county	Geographical region or administrative division
latitude	North-south position on the Earth's surface
longitude	East-west position on the Earth's surface

Table 2: Description of Destinations Columns

[2] The primary data sets that were used in this research were obtained from Kaggle and the Failte Ireland Open Data platform.Firstly The data-sets obtained from Failte Ireland Open Data Platfrom which is a website owned by the Government of Ireland which contains all the information and data-sets about all the tourist information. There we have chosen three data-sets namely called as Activities, Accommodations and Attractions.These data-sets contains various information ranging from details about the activities held, attractions and accommodation spots available in Ireland The activities data-set has about 6100 entries, whereas the accommodations and attractions data-sets contain 435 entries and the activities data-set contains 2458 entries. Also all the columns and their descriptions are shown in Table. 2, which are shared by the activity, attraction, and accommodation data-sets.
Also, the weather hourly data-set which was obtained from Kaggle is originally sourced

[2]https://www.failteireland.ie/Research-Insights/Open-data.aspx

Column	Description
county	Geographical region or administrative division
station	Weather station providing the data
latitude	North-south position on the Earth's surface
longitude	East-west position on the Earth's surface
date	Date when the weather data was recorded
rain	Amount of rainfall
temp	Temperature
wetb	Wet bulb temperature
dewpt	Dew point temperature
vappr	Vapour pressure
rhum	Relative humidity
msl	Mean sea level pressure
wdsp	Wind speed
wddir	Wind direction
sun	Sunshine duration
vis	Visibility
clht	Cloud height
clamt	Cloud amount

Table 3: Description of Weather Dataset Columns

from Met Eireann, which is the meteorological authority in Ireland. The information contained in this data set comes from 25 different locations spread out throughout 15 different counties in Ireland. This service provides hourly records beginning with the start of their record-keeping from 1989 and lasting until the end of 2018. The extensive data collection has 18 columns, each of which is labelled and is shown in Table. 3 The data set that is at the core of this research enables a detailed analysis of events, locations, housing possibilities, and weather trends.But as the data-set has values from the year 1989 so in this research we have occasionally filtered like have taken the latest possible data according to our usage in recommendation systems as well as weather prediction models[3]

3.3 Data Preprocessing

One of the major steps of our methodology is called exploratory data analysis (EDA), and it is accountable for being the pillar for an in-depth analysis of our data-sets. The EDA serves as major system for our research project that goes through the preprocessing stages, helping us in understanding the details of each data-set (which includes Activities, Attractions, Accommodations, and Weather). Then we achieve meaningful insights of our data through the use of intelligent visualisations and analytical overviews. This in-depth analysis serves as the foundation for the the next stages of preparation, which ensures that our data-sets are thoroughly studied and are well set to get into feature engineering and model implementation. Let us now go further deeper into each data-set, exploring the EDA so as to create a strong foundation for our efforts with various machine learning models.

[3]https://www.kaggle.com/datasets/conorrot/irish-weather-hourly-data/data

3.3.1 Activities, Accommodations and Attractions Datasets

The detailed view of the activities data-set, which includes 6100 rows and 9 columns, was obtained during the initial stage of Exploratory Data Analysis (EDA) for the activity data-set. In a constant effort to improve the consistency and clear information, certain columns were renamed. Like For example, "url" was changed to "URL," "AddressRegion" to "County," "AddressLocality" to "Town," "AddressCountry" to "Country," and "Tags" to "Type." Duplicate rows were then removed in order to clean up the data set. After the missing values were analysed, a small amount were identified, however as in the later stages of the project there will be an data integration in which missing values would be fully assessed. An additional processing stage was applied to the 'Type' column, which showcased the type of activity. Unique values present in the Type Column were taken out, printed, and put to the related entries. The 'Province' column was added to the data-set by introducing a mapping strategy. By accurately aligning each county with its matching province, the mapping was confirmed through the use of a table. The last stage was to save the data-set in a new CSV file after a detailed study and analysis of the data-set.
Likewise in the Accommodations data-set, we found that there are data of around 2458 rows and 9 columns. Similar to EDA process done in the Activities data-set the very same is implemented for the accommodations data-set as well and finally the last step is also the same one to save the the data-set in a new CSV file.
Finally in the Attractions data-set, we found that there are data of around 435 rows and 9 columns. Similar to the EDA process done to the above two data-sets is carried out in this dataset as well and finally the last step is also the very same process of saving the Dataset in a new CSV file. All these three saved data-sets are named respectively to be used for further processes.

3.3.2 Weather Dataset

Weather Dataset for Tourist Recommendation System:
The Weather data-set undergoes a detailed analysis during the initial phases of EDA, which clearly states that it had a considerably large amount of data, consisting of 4,660,423 rows and 18 columns. These data's include hourly data from 25 weather stations situated in 15 counties across various parts of Ireland, this data collection covers the time period of around 30 years starting from 1989 until the end of 2018. While in the process of cleaning and processing, the data from 2010 to 2018 are filtered, the dates are translated, and summary statistics are produced. The Fig. 3 states the summary statistics of the weather data-sets.Some of the other works include establishing target data points, counting data points by county, and collecting data in order to meet the total count of 10000 rows which has equal amount of data from all the counties. The resulting structured data-frame, which is saved and given the name "Cleaned Weather.csv," is used as a base for further research and the development of models.

Weather Dataset for Prediction:
We proceed with the same basic stages, which include cleaning and processing the weather data and limiting it to values of 200000. Following the completion of the summary statistics for each weather station which is shown in the above Fig. 3, we have arranged the results into a structured data frame for the main motive of doing additional analysis. Among the tasks that were carried out were the conversion of dates, the filtering of data

In [6]: `summary_df`

Out[6]:

	Station	County	Earliest Records	Latest Records	Minimum Temperature	Median Temperature	Maximum Temperature	Minimum Rain	Median Rain	Maximum Rain
19	OAK PARK	Carlow	01-apr-2004 00:00	31-oct-2019 23:00	-12.4	10.0	29.5	0.0	0.0	17.8
1	BALLYHAISE	Cavan	01-apr-2005 00:00	31-oct-2019 23:00	-15.1	9.5	28.7	0.0	0.0	41.4
22	SHANNON AIRPORT	Clare	01-apr-1990 00:00	31-oct-2019 23:00	-11.3	10.7	31.5	0.0	0.0	38.4
21	ROCHES POINT	Cork	01-apr-1990 00:00	31-oct-2019 23:00	-5.2	10.4	26.3	0.0	0.0	20.0
5	CORK AIRPORT	Cork	01-apr-1990 00:00	31-oct-2019 23:00	-7.0	9.9	27.4	0.0	0.0	22.4
15	MOORE PARK	Cork	01-apr-2004 00:00	31-oct-2019 23:00	-12.3	10.3	29.8	0.0	0.0	23.4
23	SherkinIsland	Cork	01-apr-2005 00:00	31-oct-2019 23:00	-4.8	11.1	24.6	0.0	0.0	15.4
8	FINNER	Donegal	01-apr-1999 00:00	31-oct-2019 23:00	-8.4	10.2	29.3	0.0	0.0	18.0
13	MALIN HEAD	Donegal	01-apr-1990 00:00	31-oct-2019 23:00	-4.5	10.0	25.6	0.0	0.0	16.6
3	CASEMENT	Dublin	01-apr-1990 00:00	31-oct-2019 23:00	-15.4	9.9	30.3	0.0	0.0	19.7
6	DUBLIN AIRPORT	Dublin	01-apr-1990 00:00	31-oct-2019 23:00	-11.5	9.8	28.5	0.0	0.0	26.5
20	PHOENIX PARK	Dublin	01-apr-2006 00:00	31-oct-2019 23:00	-10.4	10.0	28.3	0.0	0.0	23.9
0	ATHENRY	Galway	01-apr-2012 00:00	31-oct-2019 23:00	-6.2	9.9	29.9	0.0	0.0	22.7
12	MACE HEAD	Galway	01-apr-2005 00:00	31-oct-2019 23:00	-5.8	10.6	28.8	0.0	0.0	13.0
24	VALENTIA OBSERVATORY	Kerry	01-apr-1990 00:00	31-oct-2019 23:00	-7.7	11.1	28.2	0.0	0.0	21.4
18	NEWPORT	Mayo	01-apr-2005 00:00	31-oct-2019 23:00	-8.8	10.3	29.6	0.0	0.0	40.0
4	CLAREMORRIS	Mayo	01-apr-1990 00:00	31-oct-2019 23:00	-15.0	9.6	30.0	0.0	0.0	34.0
2	BELMULLET	Mayo	01-apr-1990 00:00	31-oct-2019 23:00	-7.3	10.5	28.2	0.0	0.0	21.2
11	KNOCK AIRPORT	Mayo	01-apr-1997 00:00	31-oct-2019 23:00	-6.9	8.7	28.5	0.0	0.0	18.7
7	DUNSANY	Meath	01-apr-2008 00:00	31-oct-2019 23:00	-13.3	9.6	27.8	0.0	0.0	32.3
16	MT DILLON	Roscommon	01-apr-2009 00:00	31-oct-2019 23:00	-16.6	9.6	30.0	0.0	0.0	25.5
14	MARKREE	Sligo	01-apr-2008 00:00	31-oct-2019 23:00	-17.3	9.6	29.4	0.0	0.0	13.2
9	GURTEEN	Tipperary	01-apr-2008 00:00	31-oct-2019 23:00	-13.4	9.7	29.7	0.0	0.0	14.9
17	MULLINGAR	Westmeath	01-apr-1990 00:00	31-oct-2019 23:00	-13.7	9.4	28.9	0.0	0.0	24.9
10	JOHNSTOWNII	Wexford	01-apr-2005 00:00	31-oct-2019 23:00	-4.5	10.2	25.6	0.0	0.0	18.6

Figure 3: Summary Statistics

for the years 2010 to 2018, and the counting of data points for each county. The next step was to create a target number of data points for each county, and then to sample the data in relevant with that motive. After the sampled data were put together into a new Data-Frame, they were one-hot encoded for categorical variables (county and station). This was done in order to prevent any errors from happening. Following the removal of features from the date-time column, the date-time column that was initially there was removed and then we found that some missing values were present in some of the columns so for that to address we used a method called imputation so basically what we did was with the help of the imputation method we calculated the mean averages of the particular column and filled the empty rows with the mean values. Then the preprocesed data is saved as new CSV file named as 'Cleaned Weather New.csv'. Also The Fig. 4, helps to understand the temperature variance in the counties present in Ireland.

3.4 Data Integration

During the crucial phase of preprocessing and data integration, the project works towards the ideal combination of various data sets in order to create an in-depth basis for

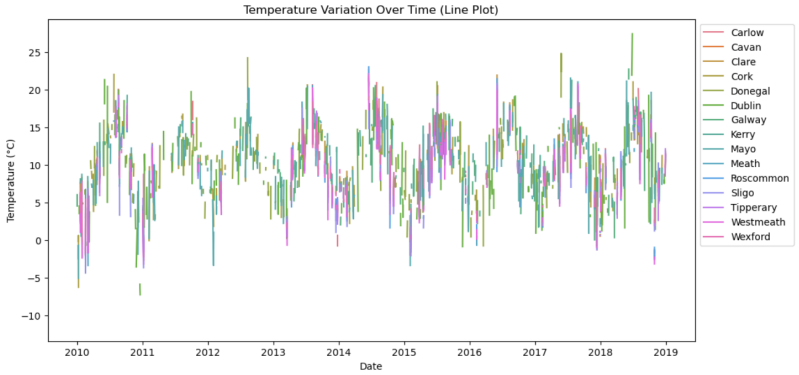

Figure 4: Temperature Variation Over Time

further analysis. The basic data sets for accommodations, attractions, and activities are concatenated along rows in order to generate a single Data-Frame that is named to as "merged df." This data-set is combination of all the mentioned three datasets and all the info is saved to a CSV file that is named as merged data-set.csv. Then a deep analysis is followed in order to find any null values that are present in the merged data set (df), and any rows that contain null values are removed. At the same time, certain columns within the 'weather' Data-Frame are converted into floating-point values, which results in the removal of string data.

```
In [329]: result
```

Out[329]:

	Name	Type	Country	County	Province	Town	Longitude	Latitude	Telephone	URL
2	Rosses Point Caravan Park (Greenlands)	Camping	Ireland	Sligo	Connacht	Rosses Point	-8.569	54.307	353719177113.0	http://www.sligocaravanandcamping.ie
3	Strandhill Caravan and Camping Park	Camping	Ireland	Sligo	Connacht	Airport Road	-8.605	54.272	353719168111.0	http://www.sligocaravanandcamping.ie
292	Seashore	B&B	Ireland	Sligo	Connacht	Ballisodare	-8.572	54.209	353719167827.0	http://www.seashoreguests.com
294	Millhouse	B&B	Ireland	Sligo	Connacht	Ballymote	-8.513	54.086	353719183449.0	http://www.sligo-accommodation.com
295	Mountain Inn	B&B	Ireland	Sligo	Connacht	Coolaney	-8.605	54.171	353719167225.0	http://www.mountaininnireland.com
...	...	...	...	...	...	...	...	...	...	...
8627	CJ's Restaurant at The Landmark Hotel	Food and Drink	Ireland	Leitrim	Connacht	Carrick-on-Shannon	-8.093	53.943	353719622222	https://www.thelandmarkhotel.com/food/private-...
8629	Bush Bar, Restaurant and Coffee Shop at The Bu...	Food and Drink	Ireland	Leitrim	Connacht	Carrick-on-Shannon	-8.094	53.946	353719671000	https://www.bushhotel.com/dining-and-bar/
8688	The Lough Ree Inn	Food and Drink	Ireland	Westmeath	Leinster	Athlone	-7.929	53.465	353906433481	https://www.facebook.com/theloughreeinn/
8889	Viking Tours	Bird Watching	Ireland	Westmeath	Leinster	Athlone	-7.942	53.423	353862621136	https://www.vikingtoursireland.ie/?utm_source=...
8974	The Shed Distillery Experience	Food and Drink	Ireland	Leitrim	Connacht	Drumshanbo	-8.044	54.047	353719317000	https://www.thesheddistillery.com/?utm_source=...

7117 rows × 25 columns

Figure 5: Integrated Dataset

In the earlier preprocessing stages we created a summary Data frame which had all the data of weather categorised by various factors. This Data frame helps us to understand the weather data scenario much better and then with that we check for null values in the

'weather' Data-Frame. Removing columns such as "sun," "vis," "clht," and "clamt" from the "weather" Data-Frame.Then with the insights obtained from Rafaat (2023) helps us getting a clear picture merge to the data-sets into a single data-set by doing the below steps i.e, Converting the latitude and longitude columns from merged df and weather to float format is the very first task, since it prepares the base for the analysis. Then to merge data-sets need to have some common columns so the common columns we choose here are the Latitude and Longitude columns. Firstly each of the taken columns have values with various decimal points so we make sure that all the decimal points in the Latitude and Longitude columns are not more than 3 and are exactly 3 so that it makes it a bit more easier to merge the data-sets. The Python library known as 'Geo-Pandas' is used to merge two GeoDataFrames, specifically gdf merged and gdf weather. The presented GeoDataFrames showcase weather information and geographical data of attractions, activities and accommodations.
As a result of the integration, geographical operations are able to be done on the combined data-set. Then it utilises the cKDTree data structure, which has been optimised for nearest-neighbor queries, to effectively get geographically nearby points. The final output, marked as the'result' DataFrame, leaves out the geometry weather and nearest idx columns, which are considered useless. This process enhances the data-set through removing unnecessary information. After a thorough integration and preprocessing process, the result.csv CSV file is used to keep the more refined data-set. The Integrated dataset is shown in Fig. 5.

3.5 Feature Engineering

The research project now proceeds with feature engineering and data-set enrichment, aiming to infuse the integrated data-set with nuanced attributes for more profound analyses.During the phase of feature engineering, certain adjustments are made to the' result' data-set with the objective of improving its overall functionality. The normalisation of the temporal dimension starts with the conversion of the 'date' column. Then, an additional feature, denoted as 'weather avg,' is added to provide the mean temperature for specific parts of the country. This geographical analysis is using the help of Geo-Pandas. Improving temporal understanding through the addition of 'year' and 'month' features. The feature 'season' is effectively created to include a seasonal feature of the data. All the features created are shown in Fig. 6.
Then **One Hot Encoding** has been performed to the Weather data-set as it had many string data's present and it wouldn't be possible to perform a model with the existing data's so we use the one hot encoding method to convert the string data into numerical data.Therefore removal of unnecessary columns as well as feature engineering and One hot encoding gives an improved data-set that is saved as result.csv. Then the customised data-set features as a solid foundation for further analysis, giving a complete view of time and geographic variations within the research project.

3.6 Modelling Approach

The modelling approach integrates content-based tourism recommendations with weather prediction to enhance the Irish tourism experience. Leveraging data-sets encompassing activities, attractions, accommodations, and weather patterns, the system tailors suggestions to individual preferences. This innovative fusion aims to provide users with person-

Out[63]:

	weather_avg	year	month	season
0	5.8	2012	1	Winter
1	5.8	2012	1	Winter
2	5.8	2012	1	Winter
3	5.8	2012	1	Winter
4	5.8	2012	1	Winter
...	...	...	...	...
7112	5.9	2011	11	Winter
7113	5.9	2011	11	Winter
7114	5.9	2011	11	Winter
7115	5.9	2011	11	Winter
7116	5.9	2011	11	Winter

7117 rows × 4 columns

Figure 6: Features

alized and weather-aware recommendations, addressing the dynamic nature of tourism in Ireland.

3.6.1 Content Based Recommendation Systems

In order to enhance the overall user experience, the integration of a content based recommendation system into this research was motivated by the study from Hamid et al. (2021) which clearly stated content based recommendation has provided better results compared to others and also with the need to provide users with customised suggestions. The method makes sure that recommendations are very much customised to the user's preferences based on the historical data by carefully analysing basic features of places, which include type, geographical location, and season. By using content-based filtering in this particular research, the system is very much able to useful insights to the user based on what their requirement is and therefore giving a better user experience.

3.6.2 Multiple Linear Regression For Weather Prediction

Multiple Linear Regression finds a balance between readability and easiness when it comes to weather prediction. By considering a linear relationship between the target variable and the independent variable, this approach allows a simple and easy analysis of the ways in which each factor influences the prediction of weather. The clear and accurate understanding of weather patterns which is the one of the objective of the research fits in with the knowledge obtained from the model's parameters; thus, Multiple

Linear Regression is a suitable approach for predicting weather patterns.

3.6.3 Random Forest Regression

The use of Random Forest regression in weather prediction is influenced by its ability to handle complex data-sets and identifying irregular correlations among weather-related factors. The ability of Random Forest to recreate complex interactions among variables such as temperature, humidity, and atmospheric pressure matches to the complex features of weather data. The purpose of this selection is to use an algorithm's adaptability to try to enhance the accuracy of predictions and to understand the complex features which control weather patterns.

4 Implementation

The research project has a dual purposed approach. The research first begin with the acquiring of data-sets from Failte Ireland Open data platform and Kaggle originally sourced from Met Eireann involving activities, attractions, accommodations, and weather records. Then the data's are visualised and are understood precisely so that we get an idea to go further on the research project. Secondly EDA is performed for all the data-sets such as finding missing values, standardisation of values and arranging the data's in a perfect order for easy understanding.
Then comes the most complex part that is the Data Integration for that to do we first concatenate all the data-sets except weather data-set as they all have same structure when it comes to their data present and then with the help of Geo pandas a library in the Python we merge all the data-sets together into one single data-set which consists of all the data's of tourist informations such as the activities, attractions and accommodations along with the weather data in the particular location with some added feature engineering. Then with this data-set we do a EDA to check for null values and for the missing values we use Data imputation method to fill all the values with their mean values and thus makes it a perfect data-set with all the info present.
The project's primary focus is the implementation of a content-based recommendation system that uses cosine similarity and TF-IDF vectorization to provide tailored recommendations based on user-specified parameters. Along with this approach, weather prediction models have been initiated to predict weather using various meteorological variables present in the dataset. Then Evaluation is done by Random Forest Regression and Multiple Linear Regression. These models are effective because they are validated by reliable evaluation measures such as Mean Squared Error ,R2 Score.
In the meantime, the research presents an in depth evaluation method by defining a function that analyses the performance of the recommendation system via Precision and Recall metrics. This dual approach highlights the project's approach to enhance visitor experiences by smart recommendations as well as offering forecasting data to handle the challenges caused by Irish tourism's seasonality.

5 Evaluation

This section provides a detailed explanation about the research performance and assessment. We'll go through the various models we used, including Random Forest Regression

and Multiple Linear Regression, and their functions. In the combination is a cool recommendation system which makes use of cosine similarity and TF-IDF. We will analyse each model, explain the reasons behind our selection, and look at the insights each one added to our research into Irish tourism. Prepare for an in-depth look at how various machine learning techniques address the issue of seasonality.

5.1 Content based Recommendation Systems

To improve the tourist's experience, we created three advanced content-based recommendation systems. These systems offer traveler-specific recommendations using cosine similarity and TF-IDF vectorization.

Cosine Similarity : Cosine similarity is a metric used to determine the cosine of the angle between two non-zero vectors. As an example In order to make recommendations that are more personalised and give better results influenced from Hamid et al. (2021) content-based filtering is being used in this research to help identify attractions that share similar features. It makes recommendations for locations based on how well their characteristics match ones the user has shown interest in.To make it simple, cosine similarity gives users a numerical measure of how similar two places or attractions are to each other based on their characteristic variables. This can improve the accuracy and importance of suggestions made by the tourism recommendation system.

TF-IDF vectorization : Textual information about tourist destinations can be examined using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorization technique, which is commonly used in natural language processing as well as data recovery. The details about each destination can be displayed as a vector, having each word providing a value based by its TF-IDF score. As a result, each destination's description is represented in a numerical representation that is unique. The recommendation system can use this vectorization for content-based filtering.For instance When a user indicates their choices or looks at a particular spot, the system can suggest other locations that have similar TF-IDF vector representations, showing textual description similarity.
Therefore with the help of above factors we have used in content based filtering to give tourists a better user recommendation where in the user gets to select the type of places wants to visit or the seasons at which they wanna go or to list out places in the particular counties so that they wanna visit. To give better understanding we have given three examples where the data is shown accordingly based on the user's preference of place, season and county.

5.1.1 Content-Based Recommendation System for Tourists By Season

In this Content based recommendation system, All the above factors are used and is developed. Then the recommendation function takes inputs like the season. It filters places based on these inputs, calculates similarity scores, and returns a list of recommended places. The example usage where we have entered "Winter" as the season and accordingly it gives the list of places to visit in Winter as shown in Fig. 7 demonstrates how to get recommendations for a specific season and the results are then saved to a CSV file.

In [28]:
```
recommendations
```

Out[28]:

	Name	Type	Town	County	Latitude	Longitude	season	weather_avg
0	Rosses Point Caravan Park (Greenlands)	Camping	Rosses Point	Sligo	54.307	-8.569	Winter	5.8
1	Strandhill Caravan and Camping Park	Camping	Airport Road	Sligo	54.272	-8.605	Winter	5.8
2	Seashore	B&B	Ballisodare	Sligo	54.209	-8.572	Winter	5.8
3	Millhouse	B&B	Ballymote	Sligo	54.086	-8.513	Winter	5.8
4	Mountain Inn	B&B	Coolaney	Sligo	54.171	-8.605	Winter	5.8
...	...	...	...	...	...	...	...	...
7112	CJ's Restaurant at The Landmark Hotel	Food and Drink	Carrick-on-Shannon	Leitrim	53.943	-8.093	Winter	5.9
7113	Bush Bar, Restaurant and Coffee Shop at The Bu...	Food and Drink	Carrick-on-Shannon	Leitrim	53.946	-8.094	Winter	5.9
7114	The Lough Ree Inn	Food and Drink	Athlone	Westmeath	53.465	-7.929	Winter	5.9
7115	Viking Tours	Bird Watching	Athlone	Westmeath	53.423	-7.942	Winter	5.9
7116	The Shed Distillery Experience	Food and Drink	Drumshanbo	Leitrim	54.047	-8.044	Winter	5.9

2813 rows × 8 columns

Figure 7: Content-Based Recommendation System for Tourists By Season

In [30]:
```
recommendations_by_type
```

Out[30]:

	Name	Type	Town	County	Latitude	Longitude	season	weather_avg
0	Rosses Point Caravan Park (Greenlands)	Camping	Rosses Point	Sligo	54.307	-8.569	Winter	5.8
1	Strandhill Caravan and Camping Park	Camping	Airport Road	Sligo	54.272	-8.605	Winter	5.8
2	Seashore	B&B	Ballisodare	Sligo	54.209	-8.572	Winter	5.8
3	Millhouse	B&B	Ballymote	Sligo	54.086	-8.513	Winter	5.8
4	Mountain Inn	B&B	Coolaney	Sligo	54.171	-8.605	Winter	5.8
...	...	...	...	...	...	...	...	...
7112	CJ's Restaurant at The Landmark Hotel	Food and Drink	Carrick-on-Shannon	Leitrim	53.943	-8.093	Winter	5.9
7113	Bush Bar, Restaurant and Coffee Shop at The Bu...	Food and Drink	Carrick-on-Shannon	Leitrim	53.946	-8.094	Winter	5.9
7114	The Lough Ree Inn	Food and Drink	Athlone	Westmeath	53.465	-7.929	Winter	5.9
7115	Viking Tours	Bird Watching	Athlone	Westmeath	53.423	-7.942	Winter	5.9
7116	The Shed Distillery Experience	Food and Drink	Drumshanbo	Leitrim	54.047	-8.044	Winter	5.9

7021 rows × 8 columns

Figure 8: Content-Based Recommendation System by Type

5.1.2 Content-Based Recommendation System by Type

Like the previous system, this one is focused on giving recommendations for locations according to a specific type, such as "hotel" or "restaurant." The filtering is carried out according to the given type, but the TF-IDF and cosine similarity calculations are the same. Here all the types have been put as an input henceforth in the Fig. 8 a list of all the suggested locations of the types are shown. Once again, this particular recommendations are saved in file as a CSV format.

5.1.3 Content-Based Recommendation System by County

In the last one Like the previous system, this one focuses on giving recommendations for places to visit according to a specific county, such as "Galway". The filtering is carried out according to the given type, but the TF-IDF and cosine similarity calculations are the same. Here all input is given as Galway for the county type and henceforth in the Fig. 9 a list of all the suggested locations of the particular county are shown. Once again, this particular recommendations are saved in file as a CSV format.

In [35]: `recommendations_by_county`

Out[35]:

	Name	Type	Town	County	Latitude	Longitude	season	weather_avg
204	Wild Atlantic Hostel	Hostel	Leenane	Galway	55.119	-7.905	Summer	14.3
916	Gullane's Hotel	Hotel	Ballinasloe	Galway	53.331	-8.221	Winter	11.9
921	Portumna House	B&B	Portumna	Galway	53.095	-8.224	Winter	11.9
925	Shearwater Hotel	Hotel	Ballinasloe	Galway	53.326	-8.220	Winter	11.9
944	Clonmoylan Self Catering Accommodation	Self Catering Accommodation	Ballyshrule	Galway	53.067	-8.315	Winter	11.9
...	...	...	...	...	...	...	...	...
6727	Curraghduff Farm	Agriculture	Oughterard	Galway	53.476	-9.410	Winter	0.5
6730	Joyce Country Sheepdogs	Visitor Farm	Leenane	Galway	53.579	-9.572	Winter	0.5
6739	Tom Sullivan - Angling Guide	Fishing	Corr Na Móna	Galway	53.513	-9.453	Winter	0.5
6742	Petersburg Outdoor Education Centre	Climbing	Clonbur	Galway	53.557	-9.385	Winter	0.5
6761	Lough Corrib Adventures	Kayaking	An Fhairche (Clonbur)	Galway	53.516	-9.405	Winter	0.5

624 rows × 8 columns

Figure 9: Content-Based Recommendation System by County

5.1.4 Evaluation Metrics for Recommendation Systems

The recommendation system's performance can be obtained from the Precision and Recall metrics, which are essential to assess it. The system's accuracy in recommending places that match precisely the user's preferences is made clear by the Precision of 1.00, which is vital for boosting user satisfaction. While there is still room for improvement, the Recall of 0.40 shows that 40 percentage of the suggested places can be obtained by the system. It's crucial to keep in mind that this result could be based on the smaller size of the data set. A larger data-set may enable the system to perform more effectively to give more types of user preferences. This understanding is important for evaluating the system's potential for development and efficiency with more data, proving it as an useful tool to enhance tourists experiences in the Irish tourism industry.

5.2 Multiple Linear Regression For Weather Prediction

In the weather prediction module of our research, we use the reliable characteristics of multiple linear regression to create a weather prediction model. The target variable, "temp," is chosen as the dependent variable (y) in this lengthy feature selection process, while key weather factors like "wetb," "dewpt," "vappr," "month," "hour," and "wdsp," are identified as the independent variables (X). This selection is done by the use of **Feature Correlation Analysis** as well as **Lasso Regression Feature Selection Technique** which selects the variables best suited for the "temp" variable which is vital for improving the predictive performance of the model.

Testing and Training: In order to check the efficiency of the model, we perform a train-test split, allocating 80 percentage of the data-set for training purposes (X train, y train), and keeping the remaining 20 percentage for testing purposes (X test, y test). By making sure the model is trained on the majority of the data and then thoroughly tested on data that hasn't been examined before, this splitting ensures an accurate assessment of the model's predictive abilities. The Multiple Linear Regression model, which is used in our predictive system, is the backbone of the scikit-learn Linear Regression class. Because of its ability with numerous independent variables, this model is a good fit for our data set, which has a number of weather factors.

$$\text{MSE}(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

$$\text{R}^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Figure 10: MSE and R Squared Formula

[4]

The model is then used to create predictions for the test set. After a thorough evaluation of the weather prediction system With the evaluation metrics such as Mean Squared Error (MSE) and R-squared (R2) as shown in score this Fig. 10 is calculated and the model obtains a score of 0.435 for MSE and obtains a R2 score of 0.982 meaning it has almost 98.2 percentage of accuracy to predict the weather by using temperature as the target variable.

5.3 Random Forest Regression For Weather Prediction

To ensure accuracy and reliability, a systematic approach is followed when using Random Forest Regression for weather prediction. The first step involves feature selection, which highly focuses on weather factors like "wetb," "dewpt," and "vappr" in addition to some of the temporal features like "month," "hour," and "wdsp" for weather prediction. This selection is done by the use of Feature Correlation Analysis which selects the variables best suited for the "temp" variable which is vital for improving the predictive performance of the model.

Metric	Value
Training Set Mean Squared Error (Random Forest Regression)	0.00406
Training Set R2 Score (Random Forest Regression)	0.9998
Test Set Mean Squared Error (Random Forest Regression)	0.00797
Test Set R2 Score (Random Forest Regression)	0.9997

Table 4: Random Forest Regression Evaluation Metrics

Testing and Training: Then the given data-set undergoes a train-test split giving 80

[4]https://hrngok.github.io/posts/metrics/

percentage of the data for training of the data and the remaining 20 percentage of the data goes for testing. The random forest regression is then performed with 100 decision trees to create a reliable model. During the training stage the model learns numerous patterns within the training data.The model is then used to create predictions for the test set. After a thorough evaluation of the weather prediction system With the evaluation metrics such as Mean Squared Error (MSE) and R-squared (R2) score this model obtains a score of 0.0236 for MSE and obtains a R2 score of 0.99 meaning it has almost 99 percentage of accuracy to predict the weather by using temperature as the target variable. The huge R2 score of 0.99 percentage shows the high accuracy of predicting the weather. Therefore to check whether this model is really accurate we evaluated it on the Test data as well as Training data as shown in Table 4 to ensure that there is no over-fitting exist and this model works perfectly.

6 Discussion

The Irish Tourism Industry is a vastly growing and is one of the big contributors to the economy of the Ireland. With that in mind the motive for the research is very clear to try to address the issues related to seasonality by creating an accurate weather prediction model along side the creation of content based recommendation system for the tourists to provide a better experience.
Here in the recommendation system the research has been framed and executed in such a way that all the given data-sets from Failte Ireland and the obtained Weather data-set from Kaggle has been integrated to make sure that whenever the user searches for a specific type of recommendation They get the recommendation along with the calculated average temperature of the particular location to give a clear insights about the place and the temperature will be at around the time of the year. Further these recommendations have been evaluated by certain evaluation metrics as well.This clearly states that this recommendation system has accurate information in terms of the available data-sets. To get even better results in terms of recommendation and evaluation metrics a large amount of data-set would be necessary to make this recommendation system even more user friendly with variety of information's to be fetched for the users and also with a large amount of data-set with more relevant features a hybrid based recommend-er system can be created to given even more recommendations with accurate results.
Then for the weather prediction model the Random Forest Regression proves to be the best model with a great accuracy of 99 percentage and It clearly states that this model has the potential to predict the temperature with the data present and also even more features in the data-set can lead to a even more reliable Weather Prediction model.

7 Conclusion and Future Work

In summary, this research has successfully come up with an in-depth tourist recommendation system which is integrated with a part for prediction of the weather and has successfully addressed the research question. With the use of content-based recommendation systems, tourists can get customised recommendations based on the historical data.The system combines activity, accommodation, and attraction data-sets with weather data-set to give a better recommendation process, giving the users a customised and context-based experience.Three content-based recommendation systems are included in the implement-

ation these systems focus towards different user queries and are categorised by county, place type, and season. The models generate suitable suggestions by using cosine similarity and TF-IDF vectorization. Furthermore, excellent accuracy and predictive ability has been shown by the weather prediction system that uses a some machine learning models which includes the Random Forest Regression model, giving a strong foundation for reliable predictions of the weather.

For Future Works and expansion of this research the real time and live data integration can make the recommendation systems more reliable and giving the users most up to date suggestions. Then acquiring up to date latest weather data and also acquiring data from even more stations can enhance the system's application of predicting the weather to a much broader range within Ireland. Finally, The tourist recommendation system could become more customer-focused and easy to use through integration with mobile devices by the development of an application.Therefore with the above research it is very evident that with the current and even up to date suggestions in the near future this research can not only act as weather prediction and recommendation system but also to boost the Irish tourism by suggesting various places and destinations to visit at different times of the year maintaining a stable tourism throughout the year.

References

Abbasi-Moud, Z., Hosseinabadi, S., Kelarestaghi, M. and Eshghi, F. (2022). Cafob: Context-aware fuzzy-ontology-based tourism recommendation system, *Expert Systems with Applications* . IF: 3.

Abbasi-Moud, Z., Vahdat-Nejad, H. and Sadri, J. (2021). Tourism recommendation system based on semantic clustering and sentiment analysis, *Expert Systems with Applications* . IF: 3.

Arif, Y. M., Nurhayati, H., Nugroho, S. M. S. and Hariadi, M. (2022). Destinations ratings based multi-criteria recommender system for indonesian halal tourism game, *International Journal of Intelligent Engineering and Systems* . IF: 3.

Dewitte, S., Cornelis, J. P., Müller, R. and Munteanu, A. (2021). Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction, *Remote Sensing* **13**(16): 3209.

dos Santos, F. N., Almeida, A. G., Martins, C., Gonçalves, R. and Martins, J. (2019). Using poi functionality and accessibility levels for delivering personalized tourism recommendations, *Computers, Environment and Urban Systems* . IF: 3.

Fathi, S., Srinivasan, R. S., Kibert, C. J., Steiner, R. L. and Demirezen, E. (2020). Ai-based campus energy use prediction for assessing the effects of climate change, *Sustainability* **12**(8): 3223.

Hamid, R. A., Albahri, A. S., Alwan, J. K., Al-qaysi, Z. T., Albahri, O. S., Zaidan, A. A., Alnoor, A., Alamoodi, A. H. and Zaidan, B. B. (2021). How smart is e-tourism? a systematic review of smart tourism recommendation system applying data management, *Computational Science Review* . IF: 3.

Jeong, C.-S., hwan Ryu, K., Lee, J.-Y. and Jung, K.-D. (2020). Deep learning-based tourism recommendation system using social network analysis. IF: 3.

Kim, N., Na, S.-I., Park, C.-W., Huh, M., Oh, J., Ha, K.-J., Cho, J. and Lee, Y.-W. (2020). An artificial intelligence approach to prediction of corn yields under extreme weather conditions using satellite and meteorological data, *Applied Sciences* **10**: 3785. **URL:** *https://www.mdpi.com/2076-3417/10/11/3785*

Li, Q., Li, S., Zhang, S., Hu, J. and Hu, J. (2019). A review of text corpus-based tourism big data mining, *Applied Sciences* . IF: 3.

Rafaat, Y. (2023). Leveraging geospatial data in python with geopandas, *KDnuggets* . **URL:** *https://www.kdnuggets.com/leveraging-geospatial-data-in-python-with-geopandas*

Wang, M. (2020). Applying internet information technology combined with deep learning to tourism collaborative recommendation system, *PLOS ONE* . IF: 3.

Wang, Z., Srinivasan, R. and Shi, J. (2016). Artificial intelligent models for improved prediction of residential space heating, *Journal of Energy Engineering* **142**(4): 04016006.

Zheng, W., Liao, Z. and Lin, Z. (2020). Navigating through the complex transport system: A heuristic approach for city tourism recommendation, *Tourism Management* . IF: 3.