**INTERIM REPORT**

**Name: ANUSHA ADEM(0792674)**

**Project Title:**

ExploreCAN, A Recommendation System for Canadian Tourists and Attractions.

**Executive Summary:**

This recommendation system is to offer customized recommendations for Canadian tourism destinations. It makes recommendations for attractions that are comparable to those that the user has already visited and appreciated using a collaborative filtering algorithm. When making recommendations, the algorithm considers several variables, including the user's tastes, geography, and seasonality.

Users will be able to quickly search for attractions, filter results by category and location, and view comprehensive information about each attraction using the system, which will be available as a smartphone application. Users can share and add attractions to their favorites list as well as bookmark them for later use.

**Goals and Objectives:**

The goals and objectives for creating the Recommendation System for Canadian Restaurants and Tourist Attractions project are as follows:

* User experience improvement: The project's primary objective is to provide customers with personalized recommendations that are relevant to their preferences and interests. The project hopes to improve user satisfaction and experience through this method.
* Enhance neighborhood businesses: By giving them more exposure and visibility to a larger audience, the recommendation system will encourage neighborhood restaurants and tourist attractions. Local businesses will benefit as a result and be able to expand and prosper.
* Boost the tourism sector: The initiative intends to promote tourism and travel within Canada by giving users accurate and individualized recommendations. The local economy will benefit from this, and it will also help the tourism sector expand.
* Improve suggestion accuracy: The program includes machine learning and artificial intelligence approaches to continually learn and enhance recommendation accuracy over time. Users will be given the most pertinent and accurate recommendations as a result.
* Provide users a variety of options: The initiative seeks to give users a variety of restaurant and tourist attraction choices, including well-known locations and undiscovered gems. Users will have a distinctive and delightful experience as a result, and local companies will be promoted.

**Challenges:**

Some of the potential challenges that could be faced during the development of a recommendation system for Canadian restaurants and tourist attractions include:

* Data quality and Collection: Collecting and preprocessing accurate data from various sources can be challenging, as data may contain inconsistencies or be incomplete. Handling the inconsistencies was also a task in the initial stage of the project, but slowly by processing the data we filtered out the required entities of the model. Initially, we have found various datasets for this recommendation system with different entities. Filtering out the data with required entities for the evaluation was a challenge in data collection.
* Algorithm selection: Choosing the most appropriate algorithm for the project can be challenging, as it depends on the data and the objectives of the project. Initially we tried different algorithms for evaluating the model but couldn’t be able to finalize it. We are still learning how to choose and optimize the most appropriate algorithm that can significantly impact the system's accuracy and performance.
* User feedback: Gathering and incorporating user feedback can be challenging, as users may have different preferences and interests that the system needs to account for.
* User privacy: Ensuring user privacy and security when collecting and processing data can be challenging, as personal information may be collected during the recommendation process.

**Business Model:**

The Recommendation System for Canadian Restaurants and Tourist Attractions might be built using different business models, including:

Freemium Model: Users could be given access to the system for free with the option to pay for premium services. Users could pay for premium features like personalized suggestions and discounts at nearby establishments, but they could access basic restaurant and attraction recommendations for free.

Commission-based model: By charging a fee to nearby restaurants and attractions for each user referral or transaction, the system might make money. This business strategy encourages neighborhood merchants to provide system users with special deals and discounts.

Subscription-based system: Users may pay a monthly or yearly charge to access the recommendation system, which may provide special benefits and savings that aren't available to non-subscribers.

Advertising model: The system may make money by charging nearby companies for advertising space that would be displayed alongside suggestions for restaurants and attractions.

Data licensing model: The system might make money by selling its data and knowledge to outside businesses like travel agencies or tourism boards.

**Cost Estimation:**

Development expenditures: These will cover the price of creating, testing, and deploying software. This expense might be in the tens of thousands to the hundreds of thousands of dollars, depending on the size and complexity of the system.

Data Collection and Integration Costs: Costs associated with data integration and collection from sources like social media, restaurant and attraction websites, and user evaluations will be included in this. The price tag on data integration and acquisition could be in the tens of thousands of dollars range.

Costs of marketing and promotion: Marketing and promotion are necessary to draw users to the system. This might cover the price of public relations, branding, and advertising. These expenses could be in the tens of thousands of dollars or less.

Costs of maintenance and upkeep: After the system is put into use, there will be ongoing expenses related to keeping it updated and relevant so that it continues to be useful. The monthly cost of these expenses could range from a few hundred to several thousand dollars.

**Project Cost Estimation Breakdown:**

**Shape

Description automatically generated with medium confidence**

**RACI Matrix:**

A tool for outlining and defining roles and responsibilities within a team or project is the RACI matrix. Responsible, Accountable, Consulted, and Informed is the abbreviation for this phrase. The Recommendation System for Canadian Restaurants and Tourist Attractions might be developed using a RACI matrix, as explained below:

Responsible: The individual or group in charge of accomplishing a given job or deliverable.

Accountable: The individual who is ultimately in charge of determining whether a task or deliverable is successful or unsuccessful.

Consulted: The individuals or groups who must offer suggestions or criticism regarding a task or deliverable.

Educated: The teams or individuals that want regular updates on the status of a task or delivery.

Diagram

Description automatically generated

**Identification and analysis of Stakeholders.**

Tourists: The system's major audience, who are searching for Canada's top eateries and tourist destinations.

Local restaurants and attractions: These are interested in having the system recommend their places to guests, and they might also be interested in advertising events and deals.

Local governments and tourism boards: These may be interested in promoting tourism in their region and making sure that the recommendation system appropriately represents the available local options.

Developers and Technical Team/Data Analysts: These parties are involved in the design, development, and upkeep of the recommendation system and are concerned with making sure it is scalable and functional.

Investors: If the recommendation system is being created as a business product, investors may be involved in offering funds for expansion and development.

Regulators: Depending on the data being used for recommendations, regulators like privacy protection authorities may be involved to make sure the data is being utilized sensibly and morally.

Chart, sunburst chart

Description automatically generated

**Individual Technical Approach:**

The objective is to give a thorough explanation of the methodology that will be used to develop a recommendation system for Canadian restaurants and tourism destinations.

ii. The system tries to give customers tailored recommendations based on their interests, previous behavior, and other pertinent factors.

iii. To create precise predictions and advance over time depending on user feedback, the system will make use of machine learning algorithms and data analysis techniques.

iv. The technical methodology for creating, testing, and deploying the recommendation system is described in this paper, along with the tools and technologies that will be used, the development procedure, and the testing procedures.

**System Architecture:**

* Data gathering: The first step includes gathering data from numerous sources, including user reviews, ratings, and location information. The recommendation system would be trained using this data to produce precise recommendations.

The data is collected from various sources such as Yelp open datasets, Trip Advisor Dataset, Canadian Tourism Commission data.

* Data cleaning and preprocessing: Once the data is collected, it would need to be cleaned and preprocessed to remove any duplicates, inconsistencies, or irrelevant data. For this, firstly the data has to be cleaned, normalized and transformed so it can be used to train the model based on the required parameters.
* Feature engineering: The data would need to be transformed into features that can be used by the recommendation system. For this, firstly we must extract the features from the data that could be helpful in making the recommendations such as the cuisine type, location, price range, population, and user ratings.
* Algorithm selection: The next step would be to select the most appropriate algorithm for generating recommendations based on the type of data and business requirement.
* Model training: The algorithm would then be trained on the preprocessed and historical data to learn the similarities, patterns, and preferences of users to generate accurate recommendations. This involves using machine learning and artificial intelligence techniques.
* Model evaluation: By Monitoring and evaluating the recommendation model system regularly and testing it with different scenarios to ensure its performance as expected. Metrics like precision, recall, and accuracy would be used to measure the performance of the model. By doing so, the model might be improved upon, and any gaps found.

* Implementation: After the model has been trained and evaluated, it will be integrated into an interface that makes it simple for users to interact and navigate the recommendation system. By Continuously improving the recommendation system by incorporating feedback from the users, adding new features, and updating the data sources.

**Data Collection:**

We have gathered information using a variety of techniques, such as user feedback, user activity tracking, social media analysis, surveys and polls, and outside sources. User profiles and algorithmic recommendations have been created using the gathered data. The following are some of the ways from where the data has been collected:

1. Web scraping: You can use web scraping tools, such as Beautiful Soup or Scrapy, to extract data from websites that list Canadian restaurants or tourist attractions, such as TripAdvisor, Yelp, or Google Maps.
2. APIs: You can use APIs provided by websites that list Canadian restaurants or tourist attractions, such as TripAdvisor or Yelp, to access their data directly. You will need to sign up for an API key and follow their API usage guidelines.
3. User-generated data: You can collect user-generated data by allowing users to rate and review Canadian restaurants or tourist attractions through your website or mobile app.
4. Public datasets: You may also find public datasets that contain information on Canadian restaurants or tourist attractions, such as open data portals or data-sharing platforms.
5. Manually collected data: You can also manually collect data by researching Canadian restaurants or tourist attractions and recording their information, such as their name, location, and rating.

**Entities for creating a Chatbot:**

1. Name of the restaurant or attraction
2. Location (address, city, province)
3. Type of cuisine or attraction (e.g., Italian restaurant, outdoor attraction)
4. User ratings and reviews
5. Price range
6. Opening and closing times
7. Amenities offered (e.g., parking, wheelchair accessibility)
8. Images or photos
9. Historical visitor data (e.g., number of visitors, peak season)
10. Popularity or trending information

**Various Types of Datasets:**

1. Yelp Open Dataset: A large dataset containing information on businesses, including Canadian restaurants and tourist attractions, as well as user reviews and ratings. The data can be accessed for free through the Yelp website.
2. TripAdvisor Dataset: A dataset containing information on tourist attractions and restaurants around the world, including Canada. The dataset includes information on location, ratings, and reviews.
3. Foursquare API: Foursquare offers an API that enables access to data on millions of establishments, including eateries and tourism attractions, by developers. The API provides location, rating, and review data.
4. Google Places API: Google offers an API that enables developers to access details about establishments, such as eateries and tourism destinations. The API provides location, rating, and review data.
5. OpenStreetMap: OpenStreetMap is a world map that is available for free and without restrictions. The map contains details on restaurants and tourist sites, such as where they are located and other information.
6. Canadian Open Data Portal: The Canadian Open Data Portal gives access to a variety of datasets on Canada, including information on travel, dining, and other related areas.

It's important to note that while these datasets can provide valuable information for your recommendation system, you may need to combine or supplement the data with other sources to ensure that you have a comprehensive and up-to-date dataset.

**Finalized Datasets:**

* 1. Candian\_resturant.xlsx
     1. Has the following entities:
        1. Restaurant Name
        2. Cuisine Type
        3. Address
        4. City
        5. Province/State
        6. Country
        7. Postal/Zip Code
        8. Phone Number
        9. Price Range (per person)
        10. Average Rating (out of 5)
        11. Number of Reviews
  2. Candian\_places.xlsx
     + 1. Longitude
       2. Latitude
       3. Name
       4. Place\_type
       5. Phone
       6. dates\_open
       7. amenities
       8. state
       9. State Name
       10. City
       11. Province/State
  3. Candian\_Provinces.xlsx
     + 1. Longitude
       2. Latitude
       3. state
       4. City
       5. Province/State

**Data Preprocessing:**

This section will outline the methods we intend to employ for pre-processing the data, such as feature selection, outlier removal, and data normalization. Additionally, it will describe the technologies and techniques we intend to use for data pre-processing.

* Data cleaning is the process of eliminating redundant or irrelevant data, filling in blanks, and fixing errors.
* Data integration is the process of merging information from various sources to produce a single dataset.
* Data transformation entails scaling data to make it more manageable or changing data from one format to another.
* The most pertinent features to include in the recommendation system must be chosen through the feature selection process.
* Data normalization entails scaling the data to fit within a predetermined range**.**

**Feature Extraction:**

The features that we intend to extract from the pre-processed data are described in this section. It will detail the feature extraction methods we intend to employ, such as sentiment analysis, text mining, and image processing.

Bag-of-Words(BoW): Natural language processing (NLP) applications frequently employ the Bag-of-Words (BoW) technique to extract features from text data. The order of the words is not important in BoW; instead, the frequency of each word is employed as a feature to represent each document or text input as a bag of words.

Collaborative Filtering: It identifies users or products that are similar based on past behaviour. User-based filtering detects similar persons based on their prior interactions, whereas item-based filtering identifies similar items based on the user's past interactions. Collaborative filtering can be either user-based or item-based.

Matrix Factorization: With the help of the matrix factorization approach, high-dimensional data can be represented in a more manageable, lower-dimensional environment. It is used in recommendation systems to discover latent features, which are not explicitly represented in the data but can be inferred from its patterns.

Deep Learning: Deep learning is a method for automatically learning hierarchical representations of the input data. Neural networks are used. In image and video recommendation systems, where the characteristics are taken from the visual material, deep learning is frequently employed.

**Recommendation Generation**

Modelling is the next step in developing a recommendation system for Canadian restaurants and tourism destinations after data preprocessing and feature extraction. Building a recommendation system for modelling entails creating user-specific recommendations using the retrieved attributes.

Collaborative filtering: It is a strategy that involves finding users or products that are similar to one another based on their prior interactions and creating suggestions based on those users' interests.

Content-Based Filtering: This method creates suggestions based on the characteristics of the products themselves, like their cuisine, price range, location, or rating.

Hybrid filtering: It blends collaborative filtering and content-based filtering to produce recommendations that are more precise and varied. Rule-based hybrid filtering creates suggestions in accordance with predetermined rules, whereas model-based hybrid filtering uses machine learning models that mix collaborative and content-based methodologies to produce recommendations.

Reinforcement Learning: With this technique, recommendations are created based on how the user interacts with the system; as a result, the system learns from the user's response and modifies the recommendations.

**User Interface:**

This section will outline the user interface that we intend to create for the system of recommendations. It will outline the design guidelines we intend to adhere to and the instruments and technologies we intend to employ to create the user interface.

Personalization: To improve the user experience, the user interface should be tailored to the user's tastes and interests. Based on the user's prior contacts with the system, such as the restaurants or tourist destinations they have visited or indicated interest in, the system should offer recommendations.

Visualization: To help the user understand the system's recommendations and offer input, the user interface should incorporate visualizations. The system can display the user's preferences and interests as well as the popularity of specific restaurants or tourist destinations via charts, graphs, or heat maps.

Interactive elements: The user interface should have interactive elements that enable feedback from the user and allow them to customize the recommendations. The system may have options like rating, review, or save buttons that let the user comment on the recommendations or save them for later use.

Seamless Integration: To improve the system's accessibility and usability, the user interface should work smoothly with the user's devices and computing platforms. The system should work with a variety of gadgets, including smartphones, tablets, and desktop computers, and it should interface with well-known platforms like social media or travel apps.

Navigation: To make it simple for people to find what they're looking for; the user interface should have clear and intuitive navigation. Menus, search bars, and filters are examples of this.

Feedback: The user interface needs to let the user know how their actions are going and, if necessary, display clear error messages.

**Machine learning Model Techniques:**

Neural Network: Deep learning architectures can be used to simulate the user-item interactions in neural networks, which can then be utilized to build recommendation systems. These algorithms can create customized suggestions by learning complex non-linear relationships between people and products.

Decision Tree: Decision trees can be used to combine collaborative filtering and content-based filtering to build a hybrid recommendation system. Decision trees can learn from user behavior and make recommendations for products depending on the user's preferences and the qualities of the products.

Support Vector Machines (SVMs): By modelling the user-item interactions as a classification problem, SVMs can be used to develop a recommendation system. The user's preference for an item can be classified by the model as either like or dislike based on how the user behaves. Then, based on the user's classification, the suggested goods are chosen.

Bayesian Network: Using Bayesian networks, a probabilistic recommendation system that offers individualized recommendations based on the user's tastes and behavior can be developed. These models can consider user comments and modify the recommendations accordingly.

Random Forest Regression: A sort of ensemble machine learning model called random forest combines different decision trees to provide a more reliable and accurate model. By anticipating the user's tastes based on their prior interactions with restaurants and tourist attractions, random forest can be utilized to develop a recommendation system for Canadian restaurants and tourist destinations.

**Random Forest Regression Machine Learning Model:**

An ensemble machine learning model called a "random forest" mixes different decision trees to produce a more reliable and accurate model. We have used this machine learning model to train the data and generate predictions. Following are the steps involved in training the model:

Data Preparation: The first stage in the process is to prepare the data by removing pertinent information and converting it into a format compatible with the random forest model. Features may include user demographics, location, previous interactions with restaurants and tourism attractions, ratings, and reviews.

Training: After the data is ready, the random forest model is trained on it to determine how the user's preferences relate to their previous interactions. The model learns from the data by building a number of decision trees, each of which is trained on a different random subset of the data and a different random subset of characteristics.

Prediction: Based on their prior interactions, the model can be used to forecast a user's preference for a restaurant or tourist destination after training. For each item, the model calculates a probability score that indicates how likely it is that the user will find the item appealing.

Recommendation: The recommended items are chosen in accordance with the likelihood scores produced by the random forest model. The user is suggested the things with the highest probability scores.

**Evaluation Metrics:**

A recommendation system's performance is measured using evaluation metrics, indicating how well it is doing in terms of giving users accurate and pertinent recommendations. We have developed a recommendation system for Canadian restaurants and tourism destinations using the following standard rating metrics:

Precision is a metric for evaluating how important the user-recommended things are. It calculates the percentage of the system's overall recommended things that are relevant items. The precision is 0.7 or 70%, for example, if the system recommends 10 items and 7 of them are relevant to the user.

Recall: Recall is a metric for how frequently the user is given recommendations for relevant items in the system. It calculates what percentage of all relevant objects the algorithm recommends are also relevant. For example, the recall is 0.7 or 70% if the system recommends 14 of the 20 relevant items.

F1-score: The F1-score combines precision and recall giving a single performance assessment of the system. It is figured out as the harmonic means of recall and precision. When there is an imbalance in the quantity of relevant and irrelevant items, the F1-score is frequently used.

Mean Absolute Error: Mean Absolute Error (MAE) is an evaluation metric that measures the average absolute difference between the predicted and actual ratings or scores. MAE can be used to gauge how well a recommendation system predicts users' ratings for suggested restaurants and tourism attractions in the context of Canadian restaurants and tourist destinations.

**Technology and Tools:**

1. Programming Languages: Python
2. Machine Learning Libraries: TensorFlow, PyTorch, Scikit-learn.
3. Big Data Processing Frameworks: Apache Spark, Hadoop
4. Database Systems: MySQL
5. Web Development Frameworks: Flask, Django, React
6. Version Control Systems: Git, SVN
7. Cloud Platforms: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)

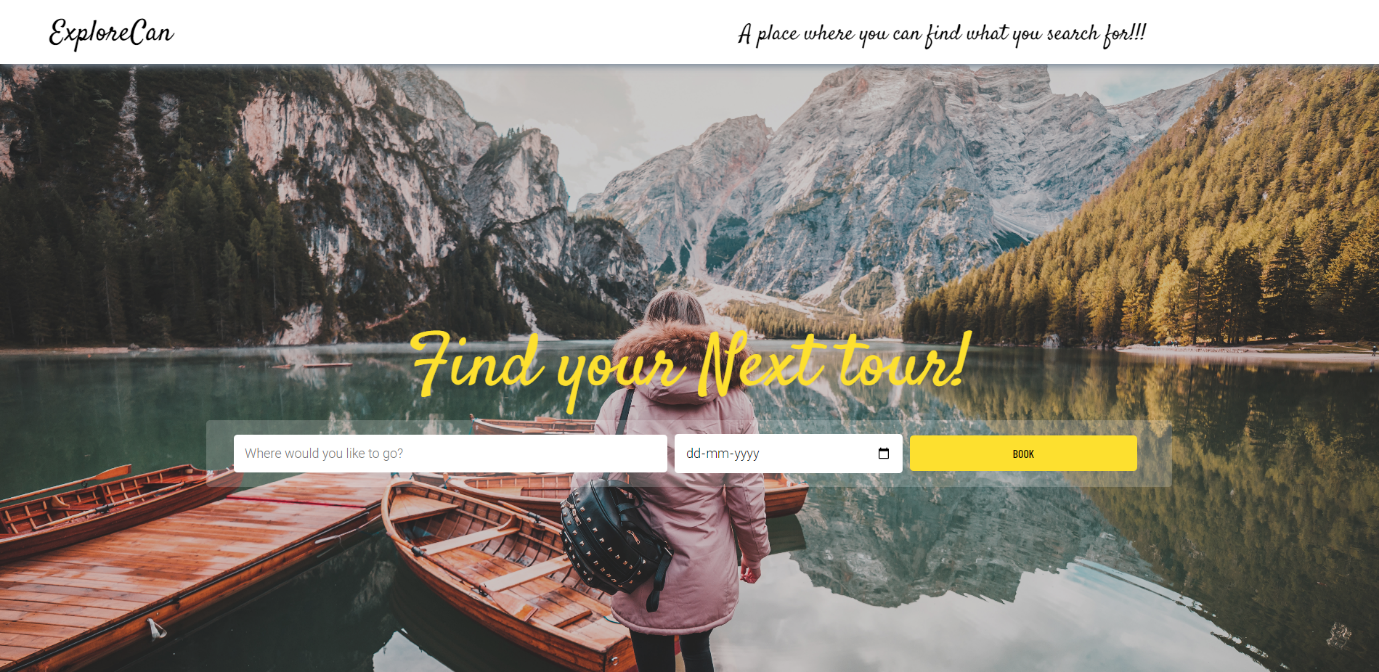
**Project Timeline:**

**Timeline

Description automatically generated**

**Current State of the Project**

User Interface:



Database:

Application, table

Description automatically generated

Graphical user interface, application

Description automatically generated

Machine Learning Model:

Table

Description automatically generated

Accuracy Testing:

Text

Description automatically generated

Graphical user interface, application

Description automatically generated

**Future Work:**

* Incorporating more data sources: The recommendation system can be improved by incorporating additional data sources, such as social media activity or location data, to provide more accurate and diverse recommendations.
* Expanding to new markets: Once the system has been successfully implemented for Canadian restaurants and tourist attractions, it can be expanded to cover other regions and countries, potentially increasing the user base and revenue.
* Collaborating with restaurants and attractions: Collaborating with restaurants and attractions to provide exclusive offers or promotions to users can help to drive business and increase user engagement with the recommendation system.

**Conclusion:**

* Improved user experience: By providing personalized recommendations based on user preferences and behavior, a recommendation system can improve the user experience for people seeking information about Canadian restaurants and tourist attractions. Users are more likely to find the information they need quickly and easily, which can increase their satisfaction and loyalty to the platform.
* Increased engagement and usage: With a recommendation system in place, users are more likely to return to the platform to discover new restaurants and attractions that align with their interests. This can lead to increased engagement and usage of the platform over time.
* Increased revenue and growth: As users discover and visit new restaurants and attractions through the recommendation system, those establishments may benefit from increased business, leading to potential revenue growth. Additionally, the platform itself may experience growth as more users are attracted by the personalized recommendations.

**References:**

1. Comprehensive Guide to build a Recommendation Engine from scratch (in Python) ***Available Online:***[*https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/*](https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/) *Accessed on: 22nd February 2023*
2. Places API ***Available Online:*** [***https://developers.google.com/maps/documentation/places/web-service***](https://developers.google.com/maps/documentation/places/web-service)*Accessed on: 12th Jan 2023*
3. Create A Travel/Tourism Website Using HTML and CSS ***Available Online:*** [***https://www.codewithrandom.com/2022/12/03/travel-website-using-html-css/***](https://www.codewithrandom.com/2022/12/03/travel-website-using-html-css/)*Accessed on: 24th Mar 2023*