Ski Resort Recommendation System Leveraging Data Mining

Juncheng Man
University of Colorado Boulder
Boulder, Colorado, USA
juma1543@colorado.edu

Audrey Ly
University of Colorado Boulder
Boulder, Colorado, USA
auly2364@colorado.edu

Grace Waida
University of Colorado Boulder
Boulder, Colorado, USA
grwa6713@colorado.edu

Justin Nguyen
University of Colorado Boulder
Boulder, Colorado, USA
jung1504@colorado.edu

1 INTRODUCTION

Skiing is one of the most popular winter sports in the United States, with a total of 60.4 million skiers reported for the 2023-24 season by The National Ski Areas Association (NSAA) [2]. As the demand for skiing continues to grow, there arises a pertinent problem to decide which ski resort within the Epic and IKON network, the two most dominant ski passes, is better and will offer the best skiing experience customised to one's preferences.

The two passes each offer access to a number of different resorts, and making a choice on the exact resort to visit is difficult as it depends on personal preferences such as location, difficulty, atmosphere, and snowfall. This problem has been compounded by the impacts of climate change, with many ski resorts reduced snowfall and seasons cutting shorter [3].

To address these challenges, we propose a ski resort recommendation system, specifically tailored to help users choose what resort they prefer within the Epic and IKON networks. Our system recommends these ski resorts based on user preferences like difficulty, location, atmosphere, and travel time, while also taking into account climate updates that can impact snowfall and skiing experience. Data mining methodologies and tools will be utilised to gain knowledge and extract patterns from past ski resorts data as well as temporal climate and snowfall data for the assembly of the recommendation system.

2 RELATED WORK

A number of previous works have focused on developing ski resort recommendation systems, predominantly in the e-commerce sector. An interactive dashboard was developed using data on ski resorts and snow coverage from the Mavens Data playground which was used for the Maven Slope Challenge [6]. The dataset were cleaned and pre-processed using Microsoft Excel, after which it would be stored in a database to be queried by users. The dashboard allows users to narrow down on their ideal ski resort using filter factors like location, season and snow, cost, slope length, and ski lift availability. However, the documentation did not provide evaluation metrics to determine the effectiveness of the solution. Another group developed a deep neural network as a recommendation system that predicts suitable ski resorts for a web agency customers [9]. The deep neural network was trained on a dataset consisting of archived data of Valraiso, a e-Commerce Web Agency of France. It takes in the features: group, budget, duration between order week and stay week, stay week, altitude, ski difficulty, size of skiing area, and country of customer, and outputs a recommended ski resort. The evaluation of the neural network show a performance with a TOP-10 accuracy that reaches 98.95% and 99.30% for French and non-French customers, respectively, where TOP-n accuracy corresponds to the accuracy where the true class matches with any one of the n most probable classes predicted by the model. The NYC Data Science Academy has also published a ski resort recommendation system developed by a student, which helps users determine the optimal ski resort using data scraped from a popular skiing website (www.onthesnow.com) [10]. Two "web spiders" were built via the Scrapy Python framework, with one extracting ski resort data and the other extracting daily snowfall data for each resort. Average annual snow totals, variance in snow totals, cumulative snowfall and lift ticket prices were used to determine the optimal location and timing for skiing. Once again, no evaluation metrics were provided to assess the effectiveness of this solution.

A common pattern across these solutions is the mining of data from resorts and snowfall datasets, and the use of popular features like cost, snowfall, and location to provide recommendations for users. However, none of them has examined in detail the role of climate in shaping the desirability of ski resorts. While the impacts of climate change on snowfall may not have been prominent when the above solutions were developed, in today's world it can no longer be ignored. Gaining knowledge from historical temporal snowfall data is no longer sufficient, and snowfall forecasts have to be combined with climate patterns to provide a more accurate recommendation for ski resorts, as we will do in our project.

3 PROPOSED WORK

3.1 Data Collection

The first step of our analysis is data collection. Our study makes use of open-source datasets from Kaggle, GitHub and government agencies. They include the Ski Resorts and Snow Coverage dataset [8] to assess snow reliability, Ticket Prices [5] to determine affordability, Type of Runs [1] to categorize resorts based on skill level, and Past Weather Data [7] to predict the likelihood of good skiing conditions. Additionally, we will look into the advantages and disadvantages of two popular ski passes, Epic vs IKON [4], to ensure that users with either pass can receive relevant recommendations. These datasets are easily downloadable from the respective websites as csv files, which can be further processed using the Python3 programming language.

3.2 Data Aggregation

We plan to aggregate data from the aforementioned sources to create a combined dataset that we will us as our source for our application. The aggregation process will involve merging data from various datasets which will allow for us to gain knowledge from the complete and big picture of skiing, and offer users comprehensive insights that are related to ski resorts. The features to aggregate will include resort features, snowfall history, ticket prices, types of ski runs, and weather or traffic data from API calls.

The aggregation process will include joining datasets based on common attributes such as resort names, locations, or time periods, ski pass accepted (IKON and Epic). This will allow us to ensure consistency across the data sources and provide a more accurate representation of the ski resorts. Additionally, we will preform data checks in subsequent steps to ensure that all datasets align correctly, with steps in place to handle missing data, duplicates, or inconsistencies between datasets

By incorporating these various datasets, we aim to provide users with the best possible recommendations, accounting for factors such as affordability, resort difficulty, snowfall, location, traffic, atmosphere, and more. This comprehensive dataset will be crucial for developing an accurate and reliable ski resort recommendation system, allowing us to tailor our suggestions based on user preferences.

3.3 Data Pre-processing

The next stage is data pre-processing, where we make sure the data is of quality and ready to be analyzed. This step involves removing duplicates, filling in any missing or incomplete information, and cleaning out any unnecessary data. Additionally, we will categorize data, like ski run types, and convert it into a format that our model can understand. Lastly, we will ensure that the information we import from our API calls is accurate and provides the necessary details. These steps will help ensure that our data is clean, accurate, and ready for analysis when we use it.

3.4 Data Warehousing

We will be using data warehousing to store the processed and aggregated datasets in a structured format, making it easy to query and analyze later. The data warehouse will act as a central repository where all the data is organized and optimized for reporting and analysis. We plan to design the schema around the structure of a star or snowflake schema to ensure all datasets are effectively queried. We will also brainstorm the best potential storage solution based on how much data we end up using, such as deciding between a relational or non-relational database and a local or cloud-based warehouse. Down the road, we will look at how to process this information quickly to retrieve data efficiently to provide recommendations to our users.

3.5 Knowledge Mining

After our data preparation and storage are ready, we will mine for valuable patterns and trends in our datasets to gain insights into the skiing landscape and understand how best to provide recommendations for end users.

Some patterns which we will look into include correlation relationship between features, such as the correlation between cost, skiing difficulty, and location with the popularity of a ski resort to gain knowledge on the most crucial factors that shape the ski resort choice of the average skier. Such correlation analysis will be performed across all sets of possible features in our dataset to extract the maximum value out of them. In addition, time-series analysis will also be performed to understand temporal evolution of some features and how relationships between our data features could change over time. This can involve exploring the historical trend of snowfall and climate conditions over time, understanding the general seasons where extremities or fluctuations can occur, and using this knowledge to perform time-series forecasting for future predictions.

In addition, we will make use of data visualization tools like Seaborn and Matplotlib in python to help us visualise the distribution of our data with respect to each other and time. This will help us to better identify outliers and trends across complex and multifaceted data like our aggregated dataset. Such visualization may also be provided to the end user for recommendation purposes.

3.6 Tools

The tools which we will adopt for the project are as follows.

(1) Python3
(2) SQL
(3) Python Cubes
(4) API Services: Google Maps API to estimate the computing distance between the users and the mountain
(5) Microsoft Excel
(6) GitHub
(7) Python Libraries: Pandas, Numpy, Matplotlib, Seaborn, Scipy, Scikit-Learn

4 EVALUATION

We will gauge user preferences, ski conditions, and factors like weather and traffic to recommend the best ski resort for each user. The calculation will include ski resort popularity, ticket prices, the type of ski runs available, and the impact of past weather conditions on the skiing experience. We will use real-time factors such as traffic conditions using the Google Maps API to determine how far the specific resort is from the user.

(1) **Accuracy of Recommendations:** The recommendation system will suggest ski resorts that align with user preferences.
(2) **Cost-Effectiveness:** The system should suggest resorts that are within the user's specified price range using ticket price data.
(3) **Real-time Accessibility:** By integrating the Google Maps API, we will consider traffic conditions and travel times using the most up-to-date information.
(4) **Ski Run Match:** The system will evaluate the different types of ski runs (green, blue, black, black diamond, etc.), and it will match the user's skill level and preferences. Success will be measured by how well the system can match these preferences based on past user data.

5 MILESTONES

During weeks one and two, we will refine our datasets to ensure they are relevant to our problem statement. We will analyze the data in weeks three through five to create graphs and begin modeling. In this phase, we will integrate data from different sources to create a cohesive dataset. We will focus on developing and implementing the system's recommendation algorithm in weeks six and seven, making sure it aligns with user's priorities like cost, location, climate factors, ski difficulty, etc. In weeks eight through ten, we will continue developing and finishing off the recommendation system based on our feedback. Finally, during Weeks 11 and 12, we will begin debugging the system to ensure its functionality and accuracy. Afterward, we will fine-tune the recommendation algorithm and continue making any adjustments needed to finalize performance. The project will conclude with a final presentation of the final system and a documentation of our journey.

6 FUTURE IDEAS

Moving forward, our group envisions expanding this project into a fully developed application that offers users a more interactive and personalized experience. The goal of the app would be to allow users to input specific preferences such as ski run difficulty, budget, and travel distance, and receive tailored ski resort recommendations that align with their needs. While we are already working on a system that generates recommendations, this enhanced app would offer more accurate suggestions and a more user-friendly interface, improving the overall experience.

Our project currently focuses on Colorado ski resorts, but we aim to expand the range of our ski resorts beyond Colorado, offering users a broader options based on their location or intended travel destinations. Additionally, we want to go beyond the Epic and IKON pass networks.

To enhance user experience further, we would integrate additional features such as user reviews, real-time weather updates, hourly snow reports, and possibly even resort-specific services like lift ticket booking. With these improvements, we believe this app has the potential to become the go-to resource for skiing enthusiasts around the world, offering everything from personalized recommendations to live updates and community feedback.

REFERENCES

[1] Ewiseman. Ski resorts stats, 2017.
[2] Forbes. U.s. ski industry had 5th busiest season despite record warm temps in 2023-24, 2024.
[3] Guardian. Ski resorts' era of plentiful snow may be over due to climate crisis, study finds., 2024.
[4] Hineline, G. Epic and ikon ski resorts data, 2022.
[5] Lai, S. Onthesnow, 2019.
[6] Malcom, O. C. Find a ski resort for your next vacation: Data analytics project, 2023.
[7] of Agriculture, D. Data from: Eleven years of mountain weather, snow, soil moisture and stream flow data from the rain-snow transition zone, 2019.
[8] Pedersen, U. T. Ski resorts, 2022.
[9] Serdouk, Y., C. T. C. E. . M. C. Ski resorts recommendation using deep neural networks. *Proceedings of the Workshop on Recommenders in Tourism Co-Located with the 15th ACM Conference on Recommender Systems (RecSys 2021) 2974* (2021), 85–89.
[10] Toujas, A. Data web scraping to help plan a ski vacation, 2017.