Proceedings of 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS 2024)
IEEE Xplore Part Number: CFP24DJ3-ART; ISBN: 979-8-3503-7999-0

AutoML based Tourism Prediction and Maximising Revenue

Prof. Sonali Mhatre
Information Technology
Bharati Vidyapeeth College of Engineering
Kharghar, India
sonalinmhatre@gmail.com

Saloni Patil
Information Technology
Bharati Vidyapeeth College of Engineering
Kharghar, India
salonipatil2801@gmail.com

Navya Mishra
Information Technology
Bharati Vidyapeeth College of Engineering
Kharghar, India
mishranavya2912@gmail.com

Vaibhav Mungelwar
Information Technology
Bharati Vidyapeeth College of Engineering
Kharghar, India
vaibhavmungelwar2001@gmail.com

Harshada Patil
Information Technology
Bharati Vidyapeeth College of Engineering
Kharghar, India
harshadasanjaypatil25@gmail.com

2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-7999-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSCSS60660.2024.10625466

***Abstract*—Tourism industry can be defined as the set of industries which facilitate the act of tourism completely by providing infrastructure and products and services and make possible travelling for different purposes and travelling to places of leisure and business interests. The main goals of tourism industry is to achieve customer satisfaction and economic development. Having these goals as a motivation, this project is built to predict whether a customer is likely to visit a place or not as well as help the tourism industry to maximize revenue. The purpose of the project is automating feature selection part of the whole ML pipeline so that manual work is reduced. The main aim of project is to analyze sales and revenue so that business side is benefitted.**

***Keywords*—AutoML, AutoKeras, Regressor, Classifier**

I. Introduction

Automated Machine Learning (AutoML) has emerged as a prominent subject within both industrial and academic realms of Artificial Intelligence (AI) research in recent times. This technology holds significant potential in offering AI solutions, particularly in regulated industries, by providing results that are explainable and reproducible. AutoML plays a crucial role in democratizing AI development, allowing individuals without extensive theoretical backgrounds in data science to participate actively.

Traditionally, each step in the data science pipeline, including data preprocessing, feature engineering, and hyperparameter optimization, necessitates manual intervention from machine learning experts. In contrast, the adoption of AutoML streamlines the development process, enabling the generation of necessary code with just a few lines. This simplification facilitates the initiation of machine learning model development.

The main focus of the project is automating feature selection part so that manual work is reduced. The main aim of project is to analyze sales and revenue so that business side is benefitted. Tourism industry can be defined as the set of industries which facilitate the act of tourism completely by providing infrastructure and products and services and make possible travelling for different purposes and travelling to places of leisure and business interests. The main goals of tourism industry is to achieve customer satisfaction and economic development. Having these goals as a motivation, this project is built to predict whether a customer is likely to visit a place or not as well as help the tourism industry to maximize revenue. It is determined by considering various factors like traffic around the area, popularity, time required and distance. The high degree of automation in AutoML aims to enable non-experts to use machine learning models and techniques without becoming machine learning experts. Additionally, automating the end-to-end machine learning application process offers the benefits of producing simpler solutions, building those solutions faster, and models that often outperform hand-designed models. In machine learning and statistics, feature selection, also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. However, this process can be time consuming if manual labour and machine learning has to be applied repeatedly.in order to solve this shortcome, we have proposed this project that aims at automating the feature selection process.

II. Literature Review

- An Automated Machine Learning (AutoML) Method of Risk Prediction for Decision Making of Autonomous Vehicles [2], studied domain specific automated machine learning for risk prediction and behavior assessment. They proposed the use of AutoML framework and Bayesian optimization.
  Merits

  - The use of AutoML frameworks can significantly reduce the time and expertise required to develop and optimize machine learning models. This is particularly beneficial for complex tasks like risk prediction in autonomous vehicles.
  - AutoML makes advanced machine learning techniques accessible to non-experts, enabling wider adoption and experimentation in various domains, including intelligent transportation system.
  - Bayesian optimization is known for its efficiency in hyper parameter tuning, often leading to better model performance compared to traditional methods such as grid search or random search.
  - Local Minima: As noted in the study, Bayesian optimization can sometimes converge to local minima rather than finding the global optimum. This can result in sub optimal model performance if the true global optimum is not reached.
  - Computational Complexity: While Bayesian optimization is efficient in many cases, it can still be computationally intensive, particularly for very large datasets or highly complex models.

- A Tourism Route-Planning Approach Based on Comprehensive Attractiveness [3] discussed planning personalized travel routes based on the perspective of tourists rather than that of tourism intermediaries. They proposed a novel rote planning method that considered multiple factors using comprehensive attractiveness index. There were observations about genetic algorithm and its efficiency. It works on the global maximum and minimum rather than focusing on local ones. Observations concluded that various factors were present in the dataset that affected the route of travelling and tourism. Efficient techniques on how to tackle travel salesman problem were discussed.
  - Personalization: The study emphasizes planning travel routes from the perspective of tourists, which enhances the personalization of travel experiences. This user-focused approach can lead to higher satisfaction and better alignment with individual preferences.
  - Comprehensive Attractiveness Index: By incorporating multiple factors into the attractiveness index, the proposed method can more accurately reflect the diverse interests and preferences of tourists, leading to more appealing and customized travel routes.
  - Global Optimization: Genetic algorithms are adept at finding global maxima and minima, avoiding the common pitfall of local optima. This can result in more efficient and optimal travel routes compared to methods that might get stuck in suboptimal solutions.
  - Real-World Applicability: The methods discussed are applicable to real-world scenarios where multiple factors, such as time, cost, and tourist preferences, need to be considered simultaneously.
  - Resource Intensive: Genetic algorithms can be computationally intensive, requiring significant processing power and time, especially for large datasets and complex route planning scenarios.
  - Parameter Sensitivity: The performance of genetic algorithms can be sensitive to their parameters (e.g., population size, mutation rate, crossover rate). Fine-tuning these parameters often requires expertise and can be time-consuming.

- Automated Feature Selection: A Reinforcement Learning Perspective [8], provided three efficient strategies to accelerate the convergence of multi-agent reinforcement learning, i.e., GMM-based generative rectified sampling strategy, rankbased SoftMax sampling strategy and IRL-based exploration strategy. It shows how to accelerate the disquisition process laterally by perfecting training strategies, and directly by perfecting disquisition strategy. The problem of automated point subspace disquisition is bandied. Through this system, we can reduce dimensionality, dock training times, enhance conception, avoid overfitting, and ameliorate prophetic delicacy in order to support downstream prophetic tasks. It's observed each point is associated to a point agent, a point agent can decide to elect or drop a point, redundancy, and applicability, and the terrain is the characteristics of the named point subspace.
  - Dimensionality Reduction: The proposed strategies for feature selection help in reducing the dimensionality of the dataset, which can lead to lower computational costs and faster training times. This is particularly valuable in handling large datasets with numerous features.
  - Avoiding Overfitting: By focusing on relevant features and eliminating redundant or irrelevant ones, the approach helps in avoiding overfitting, thus improving the generalization of the model.
  - Accelerated Convergence: The three strategies (GMM-based generative rectified sampling, rank-based SoftMax sampling, and IRL-based exploration) are designed to accelerate the convergence of multi-agent reinforcement learning. Faster convergence means that the model can reach optimal performance more quickly, saving time and computational resources.
  - Improved Exploration: The IRL-based exploration strategy enhances the model's ability to explore the feature space more effectively, leading to better feature selection and ultimately more accurate predictions.
  - The paper highlights how the approach can improve predictive accuracy by selecting the most relevant features. Better feature selection translates to models that are not only faster but also more accurate in their predictions.
  - The paper provides a systematic approach to auto-

mated feature selection, considering the environment as the characteristics of the selected feature subspace. This structured methodology ensures that each feature (or point) is evaluated in a consistent manner, enhancing the robustness of the feature selection process.
  - Local Optima: Although the strategies aim to accelerate convergence, there is still a risk of the system converging to local optima rather than finding the best global feature set. This can result in suboptimal feature selection and less accurate models.
  - Balance Between Exploration and Exploitation: Achieving the right balance between exploration (finding new features) and exploitation (using known good features) can be challenging. Overemphasis on one aspect can lead to either missing out on important features or wasting time on unimportant ones.
- AutoML for Multi-Label Classification: Overview and Empirical Evaluation [7] discussed how Automated machine literacy(AutoML) supports the algorithmic construction and data-specific customization of machine literacy channels, including the selection, combination, and parametrization of machine literacy algorithms as main ingredients. It's shown that how multilabel bracket is done and how it helps in hyperactive parameter tuning. A greedy global hunt approach grounded on hierarchical task network planning yields promising results, showing the eventuality to duly deal with the hierarchical structures that are also reflected in the model of the hunt space.
  - Algorithmic Construction and Customization: The study highlights how AutoML frameworks can support the entire pipeline of machine learning, from algorithm selection to data-specific customization. This holistic approach ensures that each step in the machine learning process is optimized for better performance.
  - Integration of Multiple Algorithms: The ability to select, combine, and parameterize different machine learning algorithms within the AutoML framework enhances flexibility and adaptability to various multi-label classification tasks.
  - Complex Problem Handling: Multi-label classification involves predicting multiple labels for each instance, which is inherently more complex than single-label classification. The study's focus on this area demonstrates the effectiveness of AutoML in handling complex classification problems.
  - Hierarchical Task Network Planning: The use of hierarchical task network (HTN) planning in the greedy global search approach is innovative. HTN planning helps in effectively navigating the search space by considering the hierarchical structures, leading to more efficient and effective search processes.
  - High Computational Demands: AutoML frameworks, particularly those involving extensive parameter tuning, can be computationally expensive. The greedy global search approach, despite its benefits, may require substantial computational resources, making it challenging to implement for large datasets or in resource-constrained environments.
  - Complex Implementation: The integration of hierarchical task network planning and the customization of machine learning pipelines can be complex to implement, requiring advanced expertise and potentially leading to increased development time.
- Tourist Prediction Using Machine Learning Algorithm [10] reviewed the techniques of machine learning that are used to predict tourism. • Also reviewed previous studies on the use of machine learning in the domain of tourism, and we used these techniques to predict number of tourists arrived in India. Studied the comparison between the prediction performances for seven algorithms, namely Decision Tree Regression, SVM, KNN, NB, SVR, Random Forest, and Logistic Regression.
  - Wide Range of Algorithms: The study reviews and compares the performance of seven different machine learning algorithms: Decision Tree Regression, SVM, KNN, NB, SVR, Random Forest, and Logistic Regression. This comprehensive comparison provides valuable insights into the strengths and weaknesses of each algorithm in the context of tourism prediction.
  - Benchmarking: By benchmarking these algorithms, the study helps identify the most effective techniques for predicting the number of tourists, aiding practitioners in selecting appropriate models for their specific needs.
  - Tourism Data Focus: The application of these machine learning techniques to predict the number of tourists arriving in India provides a practical example of how these models can be used in real-world scenarios. This focus on tourism data makes the findings directly relevant to stakeholders in the tourism industry.
  - Predictive Performance Evaluation: Evaluating the predictive performance of each algorithm on actual tourism data helps in understanding their practical utility and effectiveness, providing actionable insights for improving tourism prediction models.
  - Review of Previous Studies: The study's review of previous research on the use of machine learning in tourism offers a comprehensive overview of the current state of the field, highlighting trends, common practices, and gaps in the existing literature.
  - Identification of Effective Techniques: By analyzing and comparing various machine learning techniques, the study helps identify which algorithms are more effective for specific tasks within the domain of tourism prediction.

III. Methodology

A. User-focused

Whenever a customer or a user wants to visit any place he/she has no idea regarding the factors that determines the total expenditure to visit that place. So, the project helps to tackle this problem with the help of AutoML to achieve cost transparency thereby making the process smooth and trust worthy from user or customer perspective. AutoML helps to automate the data pre-processing to clean data and model selection for best accuracy. AutoML libraries will itself determine algorithms to pre-process data and model to be used for best accuracy for that particular dataset. In this the user has to enter which city they want to visit, age group of visitors, with whom they are travelling with, number of male and female, purpose of visit, whether they want tour arrangement, transport package, accommodation package, sightseeing package, number of nights they are planning to visit. Depending upon these 11 factors trained regression model will predict an estimated cost which will be required to visit that particular city.

B. Python Libraries Used

Regression_package_prediction: There are several regression packages that can be used for prediction in AutoML, depending on the specific use case and requirements. Some popular regression packages include: **scikit-learn:** It is a popular machine learning library in Python, which includes various regression algorithms such as Linear Regression, Ridge Regression, Lasso Regression, and Elastic Net Regression.
XGBoost: It is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It is widely used for regression problems in AutoML.
LightGBM: It is a gradient boosting framework that uses tree-based learning algorithms. It is known for its fast training speed, high accuracy, and low memory usage.
CatBoost: It is a gradient boosting library that can handle categorical features automatically and has powerful out-of-the-box performance.
These packages can be used to build regression models in AutoML frameworks such as Autokeras, H2O, TPOT, Auto-Sklearn, and Google AutoML, which automate the entire process of building and deploying models.
Tenserflow: TensorFlow is an open- source software library developed by Google for structure and planting machine literacy models. It's one of the most popular deep literacy fabrics available and provides a comprehensive ecosystem of tools and coffers for developing and training machine literacy models. Some crucial features of TensorFlow include
Flexibility: TensorFlow is largely flexible and can be used to develop a wide range of machine literacy models, including neural networks, decision trees, and retrogression models. *Scalability:* TensorFlow is designed to gauge to large datasets and complex models. It can be run on a single machine or distributed across multiple machines in a cluster.
High-level APIs: TensorFlow provides high- position APIs similar as Keras and Estimators, which make it easier to make and train models without having to write low- position law.
*Visualization tools:*TensorFlow includes tools for imaging the training process, similar as TensorBoard, which can help to identify and diagnose issues with the models.
Production readiness: TensorFlow is designed to be used in product surroundings and provides tools for planting models to a variety of platforms, including mobile bias and the web. TensorFlow supports multiple programming languages, including Python, C, Java, and Go, and can be run on a variety of tackle, including CPUs, GPUs, and TPUs(Tensor Processing Units).
AutoKeras: AutoKeras is a popular open-source AutoML framework that allows users to automatically generate and optimize deep learning models. It provides an easy-to-use interface for automating various tasks, including classification and regression. To use AutoKeras for classification or regression, follow these steps:

- Data Preprocessing: The first step is to preprocess the data. Load the data and split it into training and validation sets. Then, the data needs to be preprocessed using various techniques such as scaling, normalization, and one-hot encoding.
- AutoKeras Model Initialization: The next step is to initialize an AutoKeras model. Use ak.StructuredDataClassifier() or ak.StructuredDataRegressor() for classification and regression, respectively.
- AutoKeras Model Training: After initializing the model, train it on preprocessed data using the fit() function. AutoKeras will automatically search for the best hyper-parameters and model architecture for the problem.
- Model Evaluation: Once the model is trained, it's its performance can be evaluated on the validation set using the evaluate() function. Use the predict() function to predict the output of new data.

tenserflow.keras (load_model):tensorflow.keras is a high-level API for building and training deep learning models in TensorFlow. One of the useful features of tensorflow.keras is the ability to save and load trained models using the load_model() function.

The load_model() function can be used to load a pre-trained model saved in the HDF5 format, which is a common file format used to store Keras models.
Algorithm used
Structured_Data_Regressor
Structured data regressor is a machine learning model that is designed to predict numeric values based on input data that is structured and organized in a tabular form, such as a spreadsheet or a database. It uses regression algorithms to analyze the relationship between independent variables and dependent variable(s) and create a function that can be used to make predictions on new data. Common examples of structured data regression tasks include predicting sales, stock

prices, or the number of website visitors based on factors such as time of day, location, or advertising spend.

C. Business Focused

The project also focuses as to how a hotel can segment their customers into categories of whether or not they will cancel on the basis of the previous record of the user. By this classification the hotels can then further personalize their services to fulfil the requirements of the customers so that they carry forward with their bookings thereby maximizing their revenue. Python Libraries used:
autokeras
tenserflow
tenserflow.keras.model(load_model)
Algorithm used:
StructuredDataClassifier()
A structured data classifier is a machine learning model that categorizes information from structured data sources such as spreadsheets or databases. It works by analyzing the attributes or features of each data point and comparing them to pre-defined categories or labels.
The classifier uses algorithms such as decision trees, K-means clustering, or Naive Bayes to identify patterns in the data and make predictions about the category of each data point. The classifier is trained using a labeled dataset, where each data point is manually categorized. The model learns from these examples and adjusts its parameters to optimize its performance. In real-world applications, the classifier can be used for tasks such as document classification, product categorization, and customer segmentation. It can help organizations quickly analyze large amounts of data and make data-driven decisions more efficiently.

D. Dataset Used

User focused dataset to predict the price
Description of fields in dataset that affect final prediction.

- Travelling From: The location from which the traveler is departing for the holiday (e.g., another city, country).
- Age Group: The age group of the traveler (e.g., 18-25, 26-35, 36-45, 46-55, 55+).
- Traveling With: Information about who the traveler is journeying with (e.g., alone, partner, family, friends).
- Males: The number of male travelers in the group.
- Females: The number of female travelers in the group.
- Purpose of Visit: The reason for the trip (e.g., leisure, business, family visit).
- Tour Package:
  - Accommodation: Type of accommodation chosen (e.g., hotel, hostel, Airbnb, luxury resort).
  - Food: Level of dining preference (e.g., budget, mid-range, fine dining).
  - Activities: Types of activities planned during the holiday (e.g., sightseeing tours, adventure activities, cultural experiences).
- Number of Nights: The duration of the stay in terms of the number of nights.

With this dataset, one could analyze various factors influencing the cost of a holiday in a city, such as demographics, travel preferences, and the chosen tour package. Machine learning model is trained on this data to predict the expected cost of a holiday based on these factors.

Business focused to predict Cancellation of booking
Description of fields in dataset that affect final prediction.

- Number of Males: The count of male guests included in the booking.
- Number of Females: The count of female guests included in the booking.
- Number of Children: The count of children included in the booking.
- Week Nights: The number of nights booked for weekdays (Monday to Thursday).
- Weekend Nights: The number of nights booked for the weekend (Friday to Sunday).
- Type of Meal Plan: The type of meal plan chosen by the customer. This could include options such as:
  1) Non-AC Room with Few Amenities: A basic room without air conditioning and with limited amenities.
  2) Standard AC Room with Basic Amenities: A room equipped with air conditioning and basic facilities such as a bed, wardrobe, and attached bathroom.
  3) Deluxe AC Room with Enhanced Amenities: A more spacious room with additional amenities such as a mini-fridge, coffee maker, and work desk.
  4) Executive Suite: A premium suite offering enhanced comfort and luxury, typically featuring a separate living area, bedroom, and luxurious bathroom amenities.
  5) Presidential Suite: An opulent suite reserved for VIP guests, often featuring multiple bedrooms, a dining area, and exclusive services such as a personal butler.
  6) Penthouse Suite: A luxurious suite located on the top floor of the hotel, offering panoramic views, extravagant furnishings, and exclusive access to amenities such as a private terrace or jacuzzi.
  7) Royal Suite: The most luxurious and exclusive accommodation option, designed for discerning guests seeking the ultimate in comfort and indulgence. This suite may include lavish amenities, personalized services, and unparalleled privacy.
- Parking Space Required: Whether the customer requires parking space or not (binary: yes/no).
- Room Type: The type of room booked by the customer. This could include options such as:

1) Bed and Breakfast (B&B): This meal plan typically includes accommodation and breakfast. Guests can enjoy a variety of breakfast options provided by the hotel, such as continental breakfast, buffet breakfast, or à la carte options.
2) Half Board (HB): With the half-board meal plan, guests receive accommodation, breakfast, and one other meal, usually dinner or lunch, depending on the hotel's policy. This plan provides guests with the convenience of having at least one main meal included in their stay.
3) Full Board (FB): The full-board meal plan includes accommodation, breakfast, lunch, and dinner. Guests can enjoy a comprehensive dining experience throughout their stay, with all main meals provided by the hotel. This plan is ideal for guests who prefer the convenience of having all meals included in their booking.

- Market Segment Type: The segment the customer belongs to. This could include options such as:
  - Aviation
  - Complementary
  - Corporate
  - Offline
  - Online
- Arrival Date: The date the customer is scheduled to arrive.
- Repeated Guest: Binary indicator of whether the guest is a repeated visitor or not (yes/no).
- Previous Booking Cancellations: The number of bookings that the customer has canceled previously.
- Previous Non-canceled Bookings: The number of bookings that the customer did not cancel previously.

E. Implementation

The project is implemented as a flutter application as shown below using the mentioned algorithms. As shown in Fig 1. the project is divided into two main parts: the first one focuses on the user end and the other focuses on the business side. In the first part, the user encounters an interface which prompts the user to fill the some details for predicting the travel budget for the user for travelling to any of the six cities provided in the application. It takes into consideration various features like the number of people travelling with the user, travel package included by the user, purpose of the visitation. The second part helps the hotels to classify whether or not the user booking their services would cancel or not. For this classification it uses the previous records of the user.

User focused

It is a travel budget prediction application that requires users to input various details about their travel plans in order to get an estimated budget for their trip to one of the six cities mentioned in the app.

The features that the user needs to fill in include the age group of the traveler, who they are traveling with, the purpose

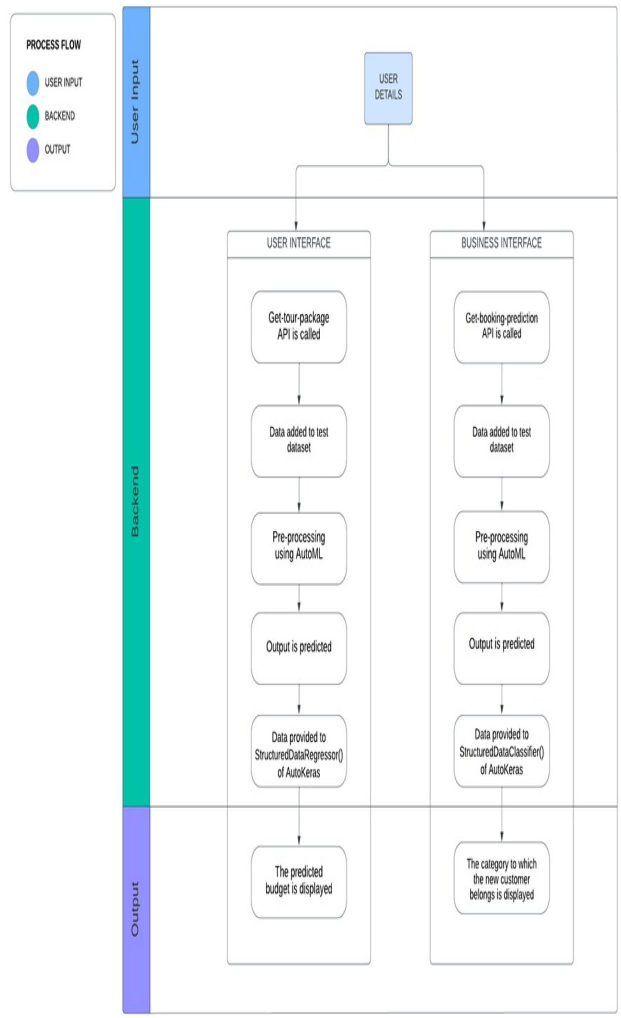

Fig. 1. Model Mean Squared Error when structured data regressor is implemented using AutoML

of the visit, the package included, the number of males and females traveling with the user, and the number of nights of stay. Based on this information, the application will use algorithms and techniques to estimate the total cost of the trip, including expenses such as accommodation, transportation, food, and other activities.

Once the user inputs all the necessary details, the predicted budget for the trip will be displayed, giving the user an idea of how much they should expect to spend on their trip. This can help users plan their trip better and make informed decisions about their travel expenses.

Business focused

Business focused part of the project uses a predictive model that classifies customers into two categories based on their booking details: cancel or will not cancel. To make this prediction, the model requires the user to input various information about the booking, including the meal plan, room type, market segment type, arrival date, average price per room, number of adults and children, weekend and week nights, previous cancelled bookings and special requests.

Based on this information, the model will use various algorithms and techniques to analyze and make a prediction about whether the customer is likely to cancel the booking or not. This prediction can be useful for hotels and other businesses in the hospitality industry to anticipate and manage their bookings and resources accordingly.

Backend

A web server is a computer program that is responsible for serving web page requests to clients. In this case, the web server is hosted on render.com, which is a cloud-based hosting platform that provides web server infrastructure for static and dynamic websites. On render.com, the web server is configured to handle HTTP requests from clients and return response data in the form of web pages or other resources.

The web server hosts two APIs that have been developed using FastAPI, a Python-based web framework for building APIs. The first API is called get-tour-package, which takes input data from the user to predict the total expenditure for a tour. This API may ask for user information such as location, duration of the tour, hotel accommodations, and transportation options. Based on the user inputs, the API uses algorithms to predict the total cost of the tour.

The second API is called get-booking-prediction, which takes customer details and helps to segment the customers. This API may ask for customer information such as meal plan, car parking space, arrival date etc. Based on the customer inputs, the API uses algorithms to segment the customers into different groups based on their preferences, behavior, and other factors.

Both APIs are connected to a Flutter application, which is a mobile app development framework for IOS and Android devices. The Flutter application uses the APIs to fetch data from the web server and display the results to the user. This means that users can access the tour package and booking prediction data from their mobile devices using the Flutter app.

Finally, the results from the APIs are displayed on the website. This means that users can access the tour package and booking prediction data from their web browsers through the website.

Outcome

Structured data regressor

The observation that the graph for AutoML shows a smooth decrease in Mean Squared Error (MSE) compared to the graph for traditional machine learning is a clear visual representation of the advantages of using AutoML in this regression task. The graphical representation provides additional insights into the performance of the two approaches:

1. AutoML's Smooth Decrease: - The smooth decrease

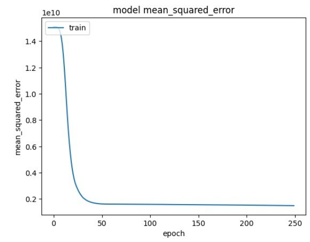

Fig. 2. Model Mean Squared Error when structured data regressor is implemented using AutoML

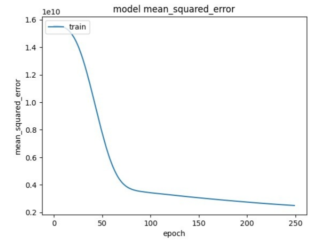

Fig. 3. Model Mean Squared Error when structured data regressor is implemented using Traditional Machine Learning

in the MSE curve for AutoML suggests that the model's performance is steadily improving over time or iterations. - A smooth, consistent decline in MSE typically indicates that the AutoML process is effectively fine-tuning the model, selecting the best algorithm, and optimizing hyperparameters.

2. Traditional Machine Learning's Graph: - In contrast, the traditional machine learning graph may show a less smooth pattern for the MSE. - This suggests that the traditional approach may struggle to converge to a better-performing model or that it requires more manual tuning and experimentation to achieve lower MSE values.

The graphical representation aligns with the previously discussed MSE values: AutoML provides a more effective and efficient solution for the regression task by consistently reducing MSE, resulting in better predictions of total costs. The traditional machine learning approach, on the other hand, may require more effort in terms of manual tuning and optimization to reach a similar level of performance.

In summary, the smooth decrease in MSE for AutoML indicates that it is a valuable tool for this regression task, automating the model selection and hyperparameter optimization processes, leading to a better-performing model. The visual representation of the MSE curves reinforces the superiority of the AutoML approach in this context.

Structured data classifier

AutoML Implementation:
Loss: 0.3149
Accuracy: 0.8760

AutoML, or Automated Machine Learning, is an approach that automates the process of model selection, hyperparameter tuning, and feature engineering. It aims to find the best-performing model for the dataset without extensive manual intervention. In this case, the AutoML implementation achieved a relatively low loss (0.3149) and a high accuracy (0.8760). These values indicate that the model built with AutoML is performing well and has a good level of accuracy in predicting whether users will cancel their bookings. The low loss suggests that the model's predictions are very close to the actual outcomes.

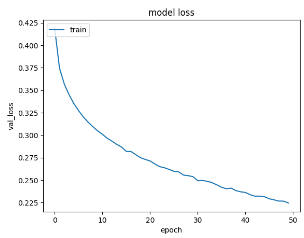

Fig. 4. Model loss when structured data classifier is implemented using AutoML

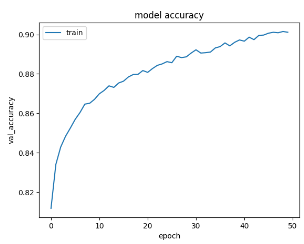

Fig. 5. Model accuracy when structured data classifier is implemented using AutoML

Traditional Machine Learning Implementation:
Loss: 0.4353
Accuracy: 0.7965
Traditional machine learning involves manually selecting algorithms, hyperparameters, and feature engineering techniques. In this approach, the loss obtained was 0.4353, and the accuracy was 0.7965. These values indicate that the traditional machine learning model also performs reasonably well but with slightly lower accuracy compared to the AutoML model. The loss is higher, and the accuracy is slightly lower, which suggests that the traditional approach may not be as optimized as the AutoML approach.

IV. Comparative Analysis

Ease of Use and Time Efficiency: AutoML simplifies the machine learning process by automating critical steps, making it accessible even to non-experts. This ease of use translates into time efficiency, as the need for manual tuning and experimentation is greatly reduced. Traditional machine learning, on the other hand, requires significant expertise and time investment to achieve comparable results.

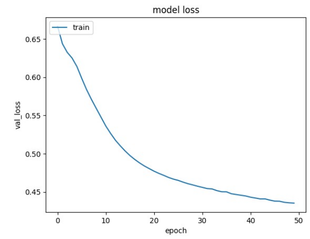

Fig. 6. Model loss when structured data classifier is implemented using Traditional Machine Learning

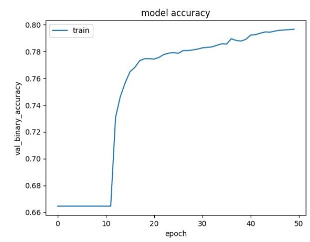

Fig. 7. Model accuracy when structured data classifier is implemented using Traditional Machine Learning

Performance and Accuracy: AutoML consistently delivers high performance due to its automated optimization processes. The smooth decrease in MSE and higher accuracy rates in classification tasks are indicative of its ability to systematically improve models. Traditional machine learning can achieve similar results but often requires extensive manual tuning and expert knowledge.

Resource Demands: While AutoML can be computationally intensive initially, the benefits of automated optimization often outweigh these demands. Traditional machine learning might have lower initial computational requirements but can become resource-intensive as manual tuning progresses.

Adaptability and Scalability: AutoML frameworks are highly adaptable, capable of handling a wide range of datasets and tasks without significant adjustments. This adaptability makes AutoML suitable for large-scale implementations. Traditional machine learning approaches vary in adaptability, often requiring custom solutions for different datasets and tasks.

Error Convergence and Stability: AutoML's automated processes typically result in a smoother and more stable convergence towards lower error rates. Traditional machine learning may experience fluctuations in error convergence, reflecting the trial-and-error nature of manual optimization.

Aspect	AutoML	Traditional Machine Learning
Model Selection	Automated	Manual
Hyperparameter Tuning	Automated	Manual
Feature Engineering	Often automated	Manual
Computational Demands	High (initially)	Variable
Ease of Use	High (user-friendly)	Requires expertise
Performance Optimization	Automated and systematic	Dependent on expertise
Resource Requirement	Higher computational resources initially	Lower initial computational demands
Adaptability	High	Variable
Time Efficiency	High (after initial setup)	Time-consuming
Scalability	High	Variable
Error/Convergence	Smooth decrease in errors	Possible inconsistent convergence
Accuracy	Generally higher	Dependent on manual tuning
Structured Data Classifier	Loss: 0.3149 Accuracy: 0.8760	Loss: 0.4353 Accuracy: 0.7965
Structured Data Regressor	Smooth and consistent decline in MSE indicates that AutoML process is effectively fine-tuning the model, selecting the best algorithm, and optimizing hyper parameters.	May struggle to converge to a better performing model or that it requires more manual tuning and experimentation to achieve lower MSE values.

TABLE I
COMPARISON BETWEEN AUTOML AND TRADITIONAL MACHINE LEARNING FOR STRUCTURED DATA TASKS.

V. CONCLUSION

The automation of the pipeline through AutoML has reduced the manual work required in the process of developing a machine learning model. This can save a significant amount of time and effort, as the manual process of selecting and optimizing machine learning models can be a time-consuming and resource-intensive task. The major benefit of AutoML that is highlighted in the project is its ability to deliver results in a shorter amount of time, leading to a speedier time to value. This can be a significant advantage in many industries, including the tourism sector, where businesses need to quickly respond to changes in customer demand and market conditions.

By automating the handling of datasets through the use of AutoML, this project can help businesses in the tourism sector maximize their profits by making accurate predictions about customer demand and optimizing their pricing strategies accordingly. The ability to automate the handling of datasets can also help to reduce the risk of errors and improve the accuracy and reliability of the machine learning model.

Overall, the automation of the pipeline through AutoML can offer significant benefits to businesses in the tourism sector and other industries by reducing the manual work required in the machine learning process and delivering results more quickly and accurately. Plus incorporating real-time data, expanding the dataset to include more cities, and providing business strategies can help to maximize the benefits of the machine learning model and support the growth and success of businesses in the tourism sector

VI. FUTURE SCOPE

The project described has several future scope opportunities for further enhancement and expansion:

Incorporating Real-time Data: Integrating real-time data streams could enhance the predictive power of the models. This could involve incorporating data from social media, weather forecasts, events happening in the area, or any other relevant sources that could impact tourism patterns and revenue generation.

Geo-spatial Analysis: Incorporating geo-spatial analysis could provide valuable insights into tourist behavior and preferences based on their location. This could involve analyzing geographical features, nearby attractions, accessibility, and other spatial factors to optimize marketing strategies and resource allocation.

Multi-modal Prediction: Expanding the prediction models to incorporate multiple modes of transportation and accommodation could provide a more comprehensive view of tourist behavior. This could involve integrating data from airlines, hotels, car rentals, etc., to offer tailored recommendations and optimize revenue across various tourism-related services.

Integration with Customer Relationship Management (CRM) Systems: Integrating the predictive models with CRM systems could enable personalized customer engagement strategies. This could involve leveraging customer data to provide tailored experiences, promotions, and loyalty programs aimed at maximizing customer satisfaction and long-term revenue.

Predictive Analytics for Seasonal Trends: Incorporate predictive analytics to forecast seasonal trends and demand fluctuations in the tourism industry. Utilize time-series forecasting models to predict peak travel periods, optimize resource allocation, and adjust pricing strategies accordingly.

REFERENCES

[1] Castelli, M., et al. (n.d.). The benefits of Automated Machine Learning in hospitality: A step-by-step guide and Automl Tool. Emerging Science Journal. https://www.ijournalse.org/index.php/ESJ/article/view/1250

[2] Shi, X., Wong, Y. D., Chai, C., and Li, M. Z. -F. (2021). An Automated Machine Learning (AutoML) Method of Risk Prediction for Decision-Making of Autonomous Vehicles. IEEE Transactions on Intelligent Transportation Systems, 22(11), 7145-7154. https://doi.org/10.1109/TITS.2020.3002419

[3] Zhang, Y., Jiao, L., Yu, Z., Lin, Z., and Gan, M. (2020). A Tourism Route-Planning Approach Based on Comprehensive Attractiveness. IEEE Access, 8, 39536-39547. https://doi.org/10.1109/ACCESS.2020.2967060

[4] Topsakal, O., Akıncı, T. C. (2023). Classification and Regression Using Automatic Machine Learning (AutoML) – Open Source Code for Quick Adaptation and Comparison. Balkan Journal of Electrical and Computer Engineering, 11(3), 257-261. https://doi.org/10.17694/bajece.1312764

[5] Chen, C., Zhang, D., Guo, B., Ma, X., Pan, G., and Wu, Z. (Year). TripPlanner: Personalized trip planning leveraging heterogeneous crowd-sourced digital footprints. IEEE Transactions on Intelligent Transportation Systems, 16(3), 1259-1273.

[6] Feurer, Matthias and Klein, Aaron and Eggensperger, Katharina and Springenberg, Jost and Blum, Manuel and Hutter, Frank, Efficient and Robust Automated Machine Learning, Advances in Neural Information Processing Systems, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2015/file/11d0e6287202fced83f79975ec59a3a6-Paper.pdf,volume28,2015.

[7] Wever M, Tornede A, Mohr F, Hullermeier E. AutoML for Multi-Label Classification: Overview and Empirical Evaluation. IEEE Trans Pattern Anal Mach Intell. 2021 Sep;43(9):3037-3054. doi: 10.1109/TPAMI.2021.3051276. Epub 2021 Aug 4. PMID: 33439834.

[8] K. Liu, Y. Fu, L. Wu, X. Li, C. Aggarwal and H. Xiong, Automated Feature Selection: A Reinforcement Learning Perspective, in IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 3, pp. 2272-2284, 1 March 2023, doi: 10.1109/TKDE.2021.3115477.

[9] Celik, Bilge, and Joaquin Vanschoren. "Adaptation strategies for automated machine learning on evolving data." IEEE transactions on pattern analysis and machine intelligence 43.9 (2021): 3067-3078.

[10] Bilal Abdualgalil Sajimon Abraham, Tourist Prediction Using Machine Learning Algorithms January 2020, Proceedings of 3rd International Conference on Sustainable Globalization - 2020

[11] Luis Sobrecueva, Automated Machine Learning with AutoKeras: Deep learning made accessible for everyone with just few lines of coding, Packt Publishing, 2021.

[12] Haifeng Jin and FranÃ§ois Chollet and Qingquan Song and Xia Hu, AutoKeras: An AutoML Library for Deep Learning, Journal of Machine Learning Research, 2023, volume 24, number 6, pages 1-6, http://jmlr.org/papers/v24/20-1355.html

[13] Kai R. Larsen, Daniel S. Becker Automated Machine Learning for Business, Book, https://books.google.co.in/books?id=d-krEAAAQBAJ&lpg=PP1&ots=ijkcz_qtly&dq=Automated%20Machine%20Learning.%20DataRobot&lr&pg=PP1#v=onepage&q&f=false

[14] Freiburg-Hannover-Tübingen, AutoML.org, https://www.automl.org/automl/