**HOTELSMART**

**Hotel Booking Cancellation Prediction using Machine Learning models**

**Name : Harsh Priyadarshan**

**Date : 01-October-2023**

**ABSTRACT:**

The "Smart Hotel Booking Assistant" is an innovative AI-driven solution designed to revolutionize the hotel booking experience for both hoteliers and travellers. Leveraging advanced machine learning algorithms and a comprehensive hotel cancellation dataset, this intelligent assistant offers a range of features to predict, optimize, and enhance the hotel reservation process.

For hotel managers, the assistant provides valuable insights into booking trends, demand forecasts, and dynamic pricing strategies. It empowers hotels to proactively manage their room inventory, reduce revenue loss due to cancellations, and enhance overall revenue performance.

With a commitment to data privacy and user experience, the Smart Hotel Booking Assistant aims to streamline hotel operations, boost profitability, and ensure travellers enjoy a seamless, stress-free booking process.

1. **PROBLEM STATEMENT:**

The Smart Hotel Booking Assistant: Our mission is to empower hotels and travellers with intelligent, data-driven solutions that optimize the hotel booking process, increase revenue, and enhancing customer satisfaction.

With the Smart Hotel Booking Assistant, we aim to:

1. **Empower Hoteliers:** We provide hotels with the tools they need to make informed decisions, reduce cancellations, and maximize revenue through dynamic pricing and demand forecasting.
2. **Delight Travelers:** We offer travellers a personalized and stress-free booking experience, tailored to their preferences, and keep them informed with real-time updates and exclusive deals.
3. **Drive Industry Efficiency:** We leverage AI and machine learning to bring efficiency to hotel operations, ensuring resources are allocated effectively, and providing valuable market insights.
4. **Ensure Data Privacy:** We prioritize data security and privacy, complying with industry regulations to protect both our customers and end-users.
5. **MARKET ASSESSMENT:**
   1. **Market Size and Growth:** Analyse the size of the hotel industry and its growth trends. Understand the market's current and projected value.
   2. **Competitive Landscape:** Identify key competitors in the hotel booking and management space. Assess their strengths, weaknesses, market share, and strategies.
   3. **Industry Trends:** Stay updated on the latest trends in the hospitality industry, such as the rise of online travel agencies (OTAs), the impact of COVID-19 on travel behaviour, and sustainability initiatives.
   4. **Regulatory Environment:** Understand the regulatory landscape, including data privacy laws, hotel industry regulations, and taxation policies that may affect the business.
6. **CUSTOMER ASSESSMENT:**

**Customer Segmentation:** Identifying and segment different types of hotel guests, such as leisure travellers, business travellers, and group bookings and understanding their preferences and behaviours.

**Customer Expectations:** Determining customer expectations when booking a hotel, including factors like price sensitivity, loyalty programs, amenities, and cancellation policies.

These problems have been the most challenging part that a based-on customer expectations what would be a chance to cancel the booking.

**BUSINESS ASSESSMENT:**

To understand the business prospect, we need to assess various factors.

1. **Revenue Optimization:** Understand the hotel's revenue goals and challenges, such as seasonal fluctuations, occupancy rates, and average daily rates (ADR).
2. **Cancellation Impact:** Analyse historical cancellation data to quantify the financial impact of cancellations on the business.
3. **Technology Infrastructure:** Evaluate the hotel's existing technology infrastructure and its compatibility with new solutions.

**Potential customers for this product include:**

1. **Hotels and Hotel Chains:** The primary customers would be individual hotels or hotel chains of various sizes. These establishments can use the assistant to optimize their room inventory, pricing, and reduce cancellations.
2. **Online Travel Agencies (OTAs):** OTAs like Booking.com, Expedia, and Agoda could benefit from integrating the assistant into their platforms to provide enhanced services to their partner hotels and improve the booking experience for travellers.
3. **Property Owners:** Owners of independent boutique hotels, bed-and-breakfasts, and vacation rentals can use the assistant to better manage their properties, increase bookings, and minimize revenue loss due to cancellations...

**4) TARGET SPECIFICATIONS AND CHARACTERIZATION**

**Hotels and Hotel Chains:**

* **Type:** Independent hotels, boutique hotels, mid-sized chains, large hotel chains.
* **Size:** Small, medium, large.
* **Location:** Urban, suburban, rural, tourist destinations.
* **Needs:** Inventory optimization, dynamic pricing, demand forecasting, guest satisfaction improvement.
* **Preferences:** Integration with existing hotel management systems, scalability, customizable features.

**Online Travel Agencies (OTAs):**

* **Type:** Global OTAs, regional OTAs, niche OTAs.
* **Services:** Accommodation booking, flight booking, vacation packages.
* **Needs:** Enhanced hotel partner services, reduced cancellation rates, improved customer retention.
* **Preferences:** Seamless API integration, real-time data exchange, competitive pricing.

**Independent Property Owners:**

* **Type:** Boutique hotels, vacation rentals, bed-and-breakfasts.
* **Ownership:** Individual owners, small property management companies.
* **Needs:** Efficient property management, increased bookings, cost-effective solutions.
* **Preferences:** User-friendly interface, affordability, customizable for small-scale operations

**5) EXTERNAL SEARCH (ONLINE INFORMATION SOURCES/REFERENCES/LINKS)**

The sources I have used as reference for analysing the problem statement Websites like Kaggle (<https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand>) and the UCI Machine Learning Repository (<https://archive.ics.uci.edu/ml/index.php>) host various datasets, including hotel-related data, which I can use for your analysis and modeling.

* 1. **BENCH MARKING ALTERNATE**

**Key Metrics and Criteria for Benchmarking:**

1. **User Experience:** Evaluate the ease of use, user interface, and overall booking experience for travelers and hoteliers.
2. **Pricing Optimization:** Compare the effectiveness of your dynamic pricing algorithms with competitors in terms of revenue optimization and reduction in cancellations.
3. **Personalization:** Analyse how well your personalized recommendation engine tailors hotel choices to individual traveller preferences.
4. **Inventory Management:** Assess how effectively your product manages room inventory and prevents overbooking compared to existing hotel management systems.
5. **Cancellation Reduction:** Measure the success of your assistant in reducing cancellations and compare it to industry benchmarks.
6. **Data Analytics:** Evaluate the depth and quality of data analytics and insights provided compared to other data analytics providers in the hospitality industry.
7. **Integration:** Consider how seamlessly your product integrates with existing hotel systems and platforms, ensuring minimal disruption during implementation.
8. **Customer Support:** Compare the level of customer support, training, and assistance you provide to hotels and travelers with that of competitors.

**Market Positioning and Unique Selling Proposition (USP):**

Identify where your product excels compared to existing products and services and use that as your Unique Selling Proposition. This could be related to predictive accuracy, affordability, user-friendliness, or any other distinctive feature that sets your assistant apart.

**Customer Feedback and Case Studies:**

Gather feedback from early adopters and conduct case studies to demonstrate the effectiveness of your product compared to alternatives. Positive customer testimonials and success stories can be powerful marketing tools.

**Continuous Improvement:**

Keep monitoring the competition and stay up to date with industry trends. Continuously improve your product based on user feedback and evolving market needs.

Benchmarking should be an ongoing process, not a one-time activity, as the competitive landscape can change rapidly in the tech and hospitality industries. Regularly reassessing your product's performance against competitors will help you maintain a competitive edge.

* 1. **APPLICABLE PATENTS**

When developing the Smart Hotel Booking Assistant, it's crucial to conduct a patent search to avoid infringing on intellectual property rights. Focus on patents related to machine learning algorithms, data analytics frameworks, booking and reservation systems, dynamic pricing, user interface elements, data privacy and security technologies, communication protocols, and blockchain if applicable. Utilize online patent databases and consider consulting a patent attorney for a comprehensive search. Conducting a patent search can be done using online patent databases, such as those provided by the United States Patent and Trademark Office (USPTO) or the European Patent Office (EPO). Collaboration with a patent attorney is advisable, particularly when dealing with complex patent matters. Stay updated on patent filings and industry trends, as new patents can emerge that may affect your product. Intellectual property rights are complex, and patent infringement can have legal and financial consequences, so it's vital to take these considerations seriously during product development.

**7) APPLICABLE REGULATIONS**

**Government Regulations:**

1. **Data Privacy and Protection:** Comply with data privacy regulations such as the European Union's General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) when handling customer data. Ensure data security and obtain necessary consents for data processing.
2. **Consumer Protection Laws:** Adhere to consumer protection laws that govern online booking, reservation, and cancellation processes. These laws vary by country and may cover issues like transparency, pricing, and cancellation policies.
3. **Taxation:** Understand the tax regulations related to the hospitality industry, including value-added tax (VAT), tourism taxes, and local occupancy taxes. Comply with tax reporting and payment requirements.
4. **Competition Laws:** Avoid anti-competitive practices that may violate antitrust or competition laws. Ensure fair and transparent pricing and business practices.
5. **Payment Processing:** Comply with financial regulations and security standards (e.g., Payment Card Industry Data Security Standard or PCI DSS) when handling payment transactions.
6. **Accessibility:** Ensure that your product and services are accessible to individuals with disabilities in accordance with accessibility laws, such as the Americans with Disabilities Act (ADA).

Top of Form

* 1. **APPLICABLE CONSTRAINTS:**

1. **Technical Constraints:**
   * **Technology Stack:** Constraints related to the choice of technology stack for development, which can impact the scalability and functionality of the product.
   * **Data Quality:** Limitations in the quality and availability of data, which can affect the accuracy of predictions and recommendations.
   * **Integration Challenges:** Challenges in integrating with various hotel management systems and online travel agencies (OTAs) due to differences in APIs and data formats.
2. **Operational Constraints:**
   * **Legal and Regulatory Compliance:** Constraints related to complying with data privacy regulations, consumer protection laws, and other legal requirements in different regions.
   * **Customer Support:** The need to provide efficient customer support to handle inquiries, complaints, and technical issues.
   * **Maintenance and Updates:** Ongoing operational requirements for system maintenance, updates, and security patches.
3. **Budgetary Constraints:**
   * **Development Costs:** Constraints related to the budget available for development, including hiring skilled developers, purchasing data sources, and software licenses.
   * **Operating Costs:** Ongoing operational costs, such as server hosting, data storage, and customer support, must align with the available budget.
4. **Time Constraints:**
   * **Development Timeline:** The need to adhere to project timelines, especially if there are market opportunities or regulatory deadlines to meet.
   * **Regulatory Deadlines:** Complying with regulatory deadlines, such as data privacy regulations, may impose time constraints on development and implementation.
5. **Resource Constraints:**
   * **Human Resources:** The availability of skilled developers, data scientists, and other experts can be constrained, particularly if there is a shortage of talent in the field.
   * **Data Availability:** The availability of comprehensive and up-to-date hotel data can be a resource constraint, as acquiring this data may require partnerships or data purchasing.
6. **Performance Constraints:**
   * **System Performance:** Constraints related to meeting performance expectations in terms of response times, recommendation accuracy, and scalability to handle peak loads.
   * **Data Processing:** The computational resources required for complex algorithms, such as dynamic pricing and prediction, may impact system performance.
7. **Space and Infrastructure Constraints:**
   * **Physical Space:** If your product requires physical infrastructure, such as data centers, consider space constraints and their impact on scalability and reliability.
   1. **FINAL PRODUCT PROTOTYPE WITH SCHEMATIC DIAGRAM**

**Data Preprocessing**

**Model deployment**

**Exploratory Data Analysis**

**Detect missing values and outliers**

**Feature Engineering**

**Model Building**

**Model deployment**

* 1. **HOW DOES AN ML SYSTEM WORK FOR HOTEL CANCELLATION?**

# Data Pre-processing:

We begin by reading in the default csv file and getting a sense of the overall size and features we will be working with.



#### Our default data file includes:

Number of columns: **32**

Number of records: **119390**

#### Imputation of missing values:

Our initial search helps us identify missing values.

A screenshot of a computer

Description automatically generatedWe will think about treatment of null values on a feature-by-feature basis. Either by fillna() imputation – using mean, median or mode or KNN imputation.

A screenshot of a computer code

Description automatically generated

**Detection of Outliers:**

The outlier detection would be done using boxplot and IQR method and will be removed in the later stages.

A screenshot of a computer

Description automatically generated

**Target Variables:**

Cancellation probability: The likelihood of a booking being cancelled.

A screenshot of a computer code

Description automatically generated

**Model Selection:**

Choosing machine learning algorithms suitable for different tasks:

* Classification algorithms- Logistic Regression, Random Foresrt, Gradient Boosting for booking and cancellation probabilities.

Our aim with this analysis is to try and understand trends in booking cancellation. Accordingly, we will build a model that will help identify which of the features are most likely to affect cancellation and consequently be useful in predicting cancellation in the future.

* As a start, we will build a logistic regression-based classifier. We shall use popular metrics such as precision and f1-score among other, to interpret the performance of our base model.

**Model Training:**

* Separating our data into target and features sections and using the Train\_Test\_Split function to create appropriate data partitions to be fed into the model.



## **Analysis and Scoring:**

Now after splitting our model in training and testing segments, and trained our model using the training segment, we shall use various methods to check performance.

After just feeding our dataset and building models using the following mentioned classifiers:

* Decision Tree
* Random Forest
* KNN

Additionally, tuning our hyperparameters for each model will be an important tool in helping improve the performance of the classifiers mentioned above.

Finally, based on the results achieved, in case there are indications of certain models being underfit, we may employ suitable techniques such as Adaboost, GradientBoost, and XGBoost. Similarly for overfit situations, we might employ methods such as regularisation, and potentially dimensionality techniques such as GLRM, and PCA more specifically.

**11) PRODUCT DETAILS AND HOW DOES IT WORK?**

**“HotelSmart”** It is a advanced management platform that combines machine learning, data analytics to enhance the hotel bookings and predicting the chances of cancellations.

**Data Sources:**

A screen shot of a computer program

Description automatically generated

Top of Form

**Algorithms, frameworks, software needed:**

**Machine Learning Algorithms:**

1. **Regression Algorithms:** Linear regression, XGBoost, or similar for pricing optimization.
2. **Classification Algorithms:** Logistic regression, Random Forest, Gradient Boosting, or neural networks for booking and cancellation probability prediction.
3. **Natural Language Processing (NLP):** NLP techniques for analyzing user reviews and comments to gauge sentiment and extract insights.

**Frameworks and Libraries:**

1. **Machine Learning Frameworks:** TensorFlow, PyTorch, or scikit-learn for building and training machine learning models.
2. **Data Processing and Analysis:** Pandas and NumPy for data manipulation and analysis.
3. **Web Development Framework:** Django, Flask, or a similar framework for building the web application.
4. **Database Management:** PostgreSQL, MySQL, or NoSQL databases like MongoDB for storing structured and unstructured data.

**Team required to develop.**

Top of Form

1. Data Scientists and Machine Learning Engineers
2. Product Manager
3. Backend and Frontend Developers
4. Business Analysts
5. Data engineers
6. Project Managers
   1. **CONCLUSION**

In conclusion, HotelSmart represents a transformative solution for the hotel industry. By harnessing advanced technologies such as machine learning, natural language processing, and data analytics, it offers travellers a personalized, hassle-free booking experience while enabling hoteliers to optimize pricing, reduce cancellations, and maximizing revenue.

HotelSmart's commitment to data privacy, ethical AI, and sustainability ensures a responsible and user-centric approach. This platform empowers both travellers and hoteliers with actionable insights, real-time updates, and transparent pricing, enhancing the overall hotel booking and management experience.