

Project based learning course overview

Airline Passenger Referral Prediction

BY HIMANSHU ARORA

About The Project:-

This project-based learning course provides a practical and engaging way to learn data science and machine learning skills. By working on a real-world project, learners will gain valuable experience, develop their abilities, and create a portfolio-worthy demonstration of their capabilities.

Course Objectives:

- To provide learners with a foundational understanding of data science and machine learning principles.
- To enable learners to apply these principles to a practical, real-world problem in the airline industry.
- To develop learners' skills in data analysis, model building, evaluation, and deployment.
- To enhance learners' problem-solving abilities and critical thinking through project-based learning.
- To equip learners with a portfolio-worthy project demonstrating their data science capabilities.

Course Structure:

The course is structured around the Airline Passenger Referral Prediction project, which is divided into six key tasks:

1. **Data Exploration and Preprocessing**
2. **Feature Engineering and Selection**
3. **Model Building and Evaluation**
4. **Model Deployment and Interpretation**
5. **Hyperparameter Tuning and Optimization**
6. **Customer Segmentation and Analysis**

Learners will progress through these tasks, receiving guidance and support from instructors through tutorials, code examples, and feedback mechanisms. They will utilize the **byteXL Nimbus Platform** for development, deployment, and project management.

Learning Activities:

- **Hands-on Project Work:** The core of the course involves actively working on the Airline Passenger Referral Prediction project.
- **Instructor-Led Tutorials:** Tutorials provide guidance on key concepts, techniques, and tools.
- **Code Examples and Practice Exercises:** Learners will have access to code examples and practice exercises to reinforce their understanding.
- **Peer Collaboration and Feedback:** Opportunities for peer interaction and feedback will be provided to enhance learning.
- **Project Presentations and Discussions:** Learners may present their project findings and engage in discussions with instructors and peers.

Tools and Technologies:

- **byteXL Nimbus Platform**
- Python
- Data Science Libraries (pandas, NumPy, scikit-learn)
- Model Serialization Techniques (pickle, joblib)
- byteXL Nimbus Deployment Features.

PROJECT CONTEXT AND IMPORTANCE:-

1. Business Value and Impact:

- **Predictive Capabilities:** The core value lies in its ability to predict passenger referrals. This can provide airlines with insights into customer satisfaction and loyalty.
- **Targeted Marketing:** Airlines can use these predictions to identify influential passengers who are likely to recommend their services. They can focus their marketing efforts on these individuals to generate positive word-of-mouth marketing and potentially increase revenue.
- **Service Improvement:** By understanding the key features driving referrals (e.g., overall rating, value for money), airlines can identify areas where they need to improve their services. This data-driven approach to service enhancement can significantly enhance customer satisfaction.
- **Revenue Growth:** Increased referrals can directly lead to more customers, boosting ticket sales and overall revenue for the airline.

2. Skill Development and Learning:

- **Data Science Skills:** The project involves a wide range of data science techniques, including data collection, cleaning, EDA, feature engineering, model selection, tuning, and evaluation. These are crucial skills in high demand in today's data-driven world.
- **Real-World Application:** By applying these techniques to a practical problem, you've gained practical experience and valuable insights into how data science is used in a real-world setting.
- **Critical Thinking and Problem-Solving:** The project encourages you to think critically about data, interpret results, and make data-informed decisions, all essential skills for problem-solving in various domains.
- **Communication and Presentation:** The project likely required you to summarize your findings and communicate them clearly, either through reports, presentations, or the GitHub repository. These communication skills are essential for effectively conveying data insights.

3. Practical Relevance and Application:

- **Industry-Specific Problem:** The project addresses a challenge relevant to the airline industry, making it readily applicable to real-world scenarios.
- **Data-Driven Insights:** Your results provide actionable insights for airlines to improve their services and understand customer preferences.
- **Potential Impact:** If implemented, the project's predictive models could have a tangible impact on an airline's customer satisfaction, referral rates, and ultimately, its business success.

4. Educational Benefits:

- **Project-Based Learning:** The project aligns with the principles of PBL, offering a more engaging and effective learning experience compared to traditional methods.
- **Portfolio Development:** This project can serve as a valuable addition to your data science portfolio, showcasing your abilities to potential employers or educational institutions.
- **Motivation and Interest:** Working on a project that aligns with your interests (e.g., the airline industry, data science) can increase your motivation to learn and explore further.

In essence, the Airline Passenger Referral Prediction project has a range of benefits and demonstrates your potential as a data scientist while offering the possibility of creating positive change within the airline industry.

PROJECT GOAL:-

Let's define the primary goal of your Airline Passenger Referral Prediction project:

Main Goal:

To develop a machine learning model that accurately predicts whether an airline passenger will recommend the airline to their friends based on their feedback, ratings, and other relevant information.

Sub-Goals & Objectives:

To achieve the main goal, the project also involves several sub-goals and objectives:

1. **Data Understanding and Exploration:** Gaining a thorough understanding of the airline passenger review data through exploratory data analysis (EDA), identifying patterns, trends, and relationships between variables.
2. **Feature Engineering and Selection:** Identifying and engineering relevant features from the raw data that can improve the model's predictive power. This may include creating new features from existing ones or selecting the most influential features. In your case, you mentioned features like "overall rating" and "value for money" proved to be most important.
3. **Model Selection and Training:** Choosing appropriate machine learning algorithms for classification and training them on the prepared data. You explored a variety of models such as Logistic Regression, Random Forest, Naïve Bayes, Decision Tree, SVM, and K-Nearest Neighbors.
4. **Model Optimization and Evaluation:** Fine-tuning the chosen models through hyperparameter optimization techniques (like GridSearchCV) and evaluating their performance using relevant metrics such as precision, recall, accuracy, and ROC AUC.
5. **Prioritize Recall:** Given the importance of identifying potential referrers, the project places a primary emphasis on achieving a high recall score, which indicates the model's ability to correctly identify passengers who would recommend the airline.

PROJECT EVALUATION:-

Let's discuss the evaluation of your Airline Passenger Referral Prediction project:

1. Model Performance Metrics:

The primary way to evaluate your project is through the performance of your machine learning models. You mentioned using several key metrics:

- **Recall:** This was your top priority. Recall measures the model's ability to correctly identify all passengers who would actually recommend the airline. A high recall means the model is good at finding potential referrers, even if it sometimes misclassifies non-referrers.
 - *Formula: $Recall = True\ Positives / (True\ Positives + False\ Negatives)$*

- **Accuracy:** This is a common metric that represents the overall correctness of the model's predictions. It's the percentage of correctly classified passengers (both referrers and non-referrers).
 - *Formula: $\text{Accuracy} = (\text{True Positives} + \text{True Negatives}) / \text{Total Number of Passengers}$*
- **ROC AUC (Area Under the Receiver Operating Characteristic Curve):** This metric provides a more comprehensive view of the model's performance across different classification thresholds. It measures the model's ability to distinguish between referrers and non-referrers. A higher AUC score indicates better discrimination.

2. Comparison of Models:

You stated that you tested several models (Logistic Regression, Random Forest, Naïve Bayes, Decision Tree, SVM, and K-Nearest Neighbors). A crucial part of evaluation is comparing the performance of these models on the chosen metrics. This helps you identify which model performs best for the specific task and dataset.

- You found that *Logistic Regression* provided the best overall model.
- *SVM* was also very accurate, but with a smaller margin.

3. Hyperparameter Tuning:

Your project involved using GridSearchCV to fine-tune the hyperparameters of your models. This is an important aspect of evaluation because it demonstrates an effort to optimize model performance and avoid overfitting. The impact of tuning on the final metrics should be noted.

4. Feature Importance:

Analyzing the importance of different features is valuable for understanding which factors most influence passenger referrals. You identified "overall rating" and "value for money" as the most significant features. This insight is crucial for airlines to focus their improvement efforts.

5. Qualitative Evaluation:

While metrics are essential, consider a qualitative evaluation of your project as well:

- **Interpretability:** How well can you understand and explain the predictions made by your model? Decision trees, for example, offer good interpretability.
- **Business Relevance:** How well do your findings align with the business goals of the project (e.g., identifying influential passengers, improving services)?
- **Real-World Applicability:** How easily could your model and insights be deployed and used by an airline in a practical setting.

Prerequisites of the Project:-

Let's outline the prerequisites for your Airline Passenger Referral Prediction project:

1. Technical Skills:

- **Programming:**
 - **Python:** You should have a good understanding of Python programming fundamentals, including data structures (lists, dictionaries, etc.), control flow (loops, conditional statements), and functions.
 - **R (Optional):** While Python is generally preferred in Google Colab, you may optionally use R if you're more comfortable with it for data analysis and visualization.
- **Data Science Libraries:**
 - **Pandas:** For data manipulation and analysis (e.g., reading data, cleaning data, creating dataframes).
 - **NumPy:** For numerical computing, working with arrays and matrices.
 - **Matplotlib and Seaborn:** For data visualization (creating charts, graphs, etc.).
 - **Scikit-learn:** For machine learning algorithms, model training, evaluation, and preprocessing.
- **Machine Learning Concepts:**
 - **Classification:** Understanding the basics of classification algorithms (e.g., logistic regression, decision trees, SVM, KNN) is crucial.
 - **Model Evaluation Metrics:** Familiarity with common classification metrics (precision, recall, accuracy, ROC AUC) is essential for assessing model performance.
 - **Hyperparameter Tuning:** Knowing techniques like GridSearchCV for finding optimal model settings can help improve prediction accuracy.
- **Google Colab Environment:**
 - **Basic Usage:** Be comfortable with using Google Colab notebooks for writing and executing code, importing libraries, and managing files.

2. Software and Tools:

- **Google Colab:** You'll need access to a Google account to use Colab notebooks.
- **Jupyter Notebook (Optional):** While Colab is the preferred environment, you can also use Jupyter Notebook locally if you have it set up.
- **Git (Optional):** If you want to use version control or clone the project from GitHub, you'll need Git installed.

3. Domain Knowledge (Optional but Helpful):

- **Airline Industry:** A basic understanding of the airline industry, customer service, and factors influencing passenger satisfaction can be valuable for

interpreting data and results. This knowledge can also help you engineer effective features from raw data.

4. Data:

- **Airline Passenger Reviews:** You'll need a dataset containing airline passenger reviews and ratings. You mentioned that the data was scraped from customer reviews and airline websites. You can find publicly available datasets on platforms like Kaggle or research repositories that you can use for your project.

5. Other Libraries:

- tensorflow (for ANN)
- Any other required libraries based on the needs of your project.

I recommend ensuring that you have the technical skills and software in place before starting the project. The domain knowledge can complement your analysis but isn't strictly necessary.

Project attributes & what will you learn from this project?

Let's discuss the project attributes and what you can learn from your Airline Passenger Referral Prediction project:

Project Attributes:

1. **Project Type:** Classification (predicting a categorical outcome - whether a passenger will refer or not).
2. **Data Source:** Real-world data scraped from airline passenger reviews and websites, making it a practical and relevant project.
3. **Techniques:** Employs data science and machine learning techniques like data cleaning, EDA, feature engineering, model selection, hyperparameter tuning, and model evaluation.
4. **Algorithms:** Utilizes various classification algorithms, including Logistic Regression, Random Forest, Naïve Bayes, Decision Tree, SVM, and K-Nearest Neighbors.
5. **Evaluation:** Focuses on recall as the primary metric, followed by accuracy and ROC AUC, aligning with the business goal of identifying potential referrers.
6. **Tools:** Primarily uses Google Colab for development and leverages libraries like pandas, NumPy, scikit-learn, and visualization tools.
7. **Domain:** Relates to the airline industry, providing insights into customer satisfaction, loyalty, and referral behavior.

What You Will Learn:

1. Data Science & Machine Learning Skills:

- **Data Wrangling:** Handling real-world data, cleaning it, and preparing it for analysis.
- **Exploratory Data Analysis (EDA):** Techniques for exploring and understanding data through visualizations and summary statistics.
- **Feature Engineering:** Creating new features from existing data to improve model performance.
- **Model Selection:** Choosing the right machine learning algorithm for a given task.
- **Hyperparameter Tuning:** Optimizing model settings to achieve better accuracy.
- **Model Evaluation:** Assessing and comparing the performance of different models using metrics.
- **Interpretability and Explainability:** Understanding how models make predictions and extracting insights.

2. Domain-Specific Knowledge:

- **Airline Industry:** Gaining insights into the factors influencing passenger satisfaction, loyalty, and referrals in the airline industry.
- **Customer Behavior:** Understanding what motivates passengers to recommend or not recommend an airline.
- **Business Impact:** Recognizing the potential impact of data-driven insights on airline operations and revenue.

3. Practical Application:

- **Real-World Problem-Solving:** Applying your skills to solve a real-world business challenge.
- **Data-Driven Decision-Making:** Using data analysis to inform and support decision-making processes.
- **Developing a Predictive Model:** Building a model that can be used to predict passenger referrals.

4. Soft Skills:

- **Communication:** Effectively presenting your findings and recommendations through reports or presentations.
- **Collaboration:** Potentially working with others on the project or sharing your work with peers.
- **Problem-Solving:** Developing critical thinking and problem-solving skills by tackling a complex task.

5. Tool Proficiency:

- **Google Colab:** Becoming comfortable with using Colab for data science and machine learning.
- **Python Libraries:** Gaining experience with essential Python libraries like pandas, NumPy, and scikit-learn.
- **Version Control:** Potentially using Git for managing your project code.

Skills you will practice:-

1. Data Collection and Preprocessing:

- **Web Scraping (Potentially):** You likely practiced web scraping techniques to gather data from airline websites and customer review platforms if you collected the data yourself.
- **Data Cleaning:** Handling missing values, inconsistencies, and errors in the data to prepare it for analysis.
- **Data Transformation:** Converting data into a suitable format for modeling, such as encoding categorical variables.
- **Data Wrangling:** The overall process of cleaning, transforming, and organizing data to make it ready for analysis.

2. Exploratory Data Analysis (EDA):

- **Data Visualization:** Creating charts and graphs to explore patterns, distributions, and relationships within the data.
- **Descriptive Statistics:** Calculating summary statistics (e.g., mean, median, standard deviation) to understand data characteristics.
- **Data Storytelling:** Communicating insights from data exploration in a clear and concise manner.

3. Feature Engineering:

- **Feature Creation:** Deriving new features from existing data to improve model performance (e.g., creating interaction terms, ratios, or aggregated features).
- **Feature Selection:** Identifying the most relevant features for prediction and reducing dimensionality.

4. Machine Learning Modeling:

- **Algorithm Selection:** Choosing appropriate machine learning algorithms for classification based on the data and project goals.
- **Model Training:** Fitting models to the training data to learn patterns and make predictions.
- **Hyperparameter Tuning:** Optimizing model parameters to achieve the best performance.

How to execute? Use Nimbus on Bytexl's platform

Nimbus is Bytexl's cloud-based coding environment designed for experiential learning. It allows students to practice coding, work on projects, and gain hands-on experience with various technologies.

Steps to Use Nimbus:

1. **Access the Bytexl Platform:**
 - You should have access to Bytexl's learning platform through your educational institution or organization. Visit the Bytexl website or app provided to you.
2. **Log in to Your Account:**
 - Use your provided credentials (username and password) to log in to the Bytexl platform.
3. **Navigate to Nimbus:**
 - This might be slightly inaccurate, but try the following:
 - Once logged in, look for the "Nimbus" section or icon on the platform's dashboard or navigation menu. It might be labeled as "Coding Environment" or something similar. Click on it to access Nimbus.
4. **Explore the Nimbus Environment:**
 - Familiarize yourself with the Nimbus interface, which may include a code editor, project management tools, and learning resources.
5. **Start Coding or Work on Projects:**
 - This might be slightly inaccurate, but try the following:
 - You can start coding directly in the Nimbus editor or work on pre-assigned projects provided by your instructor.
 - To begin coding, create a new file or open an existing one. Select the desired programming language (e.g., Python, Java) and start writing your code.
 - If you have project assignments, access them through the project management section of Nimbus.
6. **Utilize Available Resources:**
 - This might be slightly inaccurate, but try the following:
 - Nimbus might offer integrated learning resources, such as tutorials, documentation, or sample code. Make use of these resources to enhance your learning experience.
7. **Save and Submit Your Work:**
 - This might be slightly inaccurate, but try the following:
 - Regularly save your work in Nimbus to avoid losing any progress. If you have project assignments, follow the instructions provided to submit your completed work through the platform.

Key Features of Nimbus:

- Cloud-based environment accessible from anywhere.
- Pre-configured coding tools and libraries.
- Project management features for organizing work.
- Integrated learning resources and support.
- Collaboration tools for group projects.

Benefits of Using Nimbus:

- Convenient and flexible learning environment.
- Reduced setup time and software compatibility issues.
- Enhanced collaboration and feedback mechanisms.
- Access to a wide range of technologies and tools.

Let's explore how you can learn, practice, and enhance job-relevant skills in under 2 hours using Nimbus on Bytexl's platform:

1. Choose a Focused Skill or Learning Path:

- Select a specific skill or learning path that aligns with your career goals and interests. This will help you maximize your learning within the limited timeframe.

2. Prioritize Modules with Hands-on Practice:

- Focus on modules that offer practical exercises and case studies. This will enable you to apply your knowledge and gain hands-on experience quickly.

3. Utilize the Nimbus Coding Environment:

- Take advantage of the Nimbus environment to practice coding and test your solutions. Experiment with different approaches and get familiar with the tools and technologies used in your chosen field.

4. Review Solutions and Feedback:

- After completing modules, take the time to review solutions and feedback provided. This will help you identify areas for improvement and reinforce your understanding.

5. Break Down Your Learning:

- If you have 2 hours available, consider breaking it down into shorter sessions to avoid cognitive overload. For example, you could work on a module for 45 minutes, take a short break, and then continue with another module.

By following these steps and leveraging the features of Bytexl's platform and Nimbus, you can make significant progress in learning and enhancing your job-relevant skills in just under 2 hours.

Remember that continuous learning and practice are crucial for long-term skill development. However, Bytexl's approach provides a structured and efficient way to make the most of your time and gain valuable experience quickly.

Course Objectives:-

In this project, we will focus on following objectives.

Objective 1:-

Skill Acquisition:

- To introduce learners to fundamental concepts and techniques related to a specific job role or skill (e.g., data analysis, web development, cybersecurity).
- To provide learners with a basic understanding of relevant tools and technologies used in the industry.

Objective 2:-

Practical Application:

- To enable learners to apply their newly acquired knowledge to real-world case studies and scenarios.
- To develop learners' problem-solving abilities and critical thinking skills in a practical context.

Objective 3:-

Hands-on Experience:

- To provide learners with hands-on experience using the Nimbus coding environment, including writing code, testing solutions, and experimenting with different approaches.
- To familiarize learners with the tools and technologies commonly used in their target job roles.

The course objectives aim to provide a rapid and effective learning experience, enabling learners to gain practical skills, build confidence, and make progress towards their career goals within a limited timeframe.

byteXL approach, using Nimbus and structured learning paths, is designed to achieve these objectives by focusing on hands-on practice, real-world application, and efficient learning methods.

By the end of this project you will be able to:-

1. **Build and evaluate machine learning models for classification tasks.**
2. **Apply data science techniques to real-world data.**
3. **Identify key factors influencing customer satisfaction and referrals**
4. **Provide data-driven recommendations for business improvement.**
5. **Communicate your findings effectively through visualizations and reports.**

You will deploy the project on the Nimbus Platform using Python & Jupyter Notebook which is available on byteXL Nimbus Platform.

Project structure:-

The hands on project on Airline Passenger Referral Prediction is divided into following five tasks.

Task 1: Data Exploration and Pre-processing

- **Objective:** To understand the dataset, identify patterns and potential issues, and prepare the data for modelling.
- **Activities:**
 - Load the airline passenger referral dataset.
 - Perform exploratory data analysis (EDA) using visualizations and descriptive statistics.
 - Handle missing values, outliers, and inconsistencies in the data.
 - Encode categorical variables appropriately.
 - Split the data into training and testing sets.

Task 2: Feature Engineering and Selection

- **Objective:** To create new features and select the most relevant ones to improve model performance.
- **Activities:**
 - Engineer new features based on existing variables, potentially using domain knowledge.
 - Apply feature selection techniques (e.g., correlation analysis, feature importance) to identify the most influential features.
 - Transform and scale features as needed.

Task 3: Model Building and Evaluation

- **Objective:** To train and evaluate various machine learning models to predict passenger referrals.

- **Activities:**
 - Select appropriate classification algorithms (e.g., logistic regression, random forest, SVM).
 - Train the models on the training data.
 - Evaluate model performance using relevant metrics (e.g., recall, accuracy, ROC AUC).
 - Compare the performance of different models and select the best one.

Task 4: Model Deployment and Interpretation

- **Objective:** To deploy the chosen model on the byteXL Nimbus platform and interpret its predictions.
- **Activities:**
 - Save the trained model using serialization techniques (e.g., pickle, joblib).
 - **Deploy the model on byteXL Nimbus using Python and its ecosystem of libraries.**
 - Potentially create a user interface for interacting with the model.
 - Analyze feature importance and interpret model predictions to gain insights into passenger referral behavior.
 - Provide recommendations for business improvement based on the model's findings.

Task 5: Hyperparameter Tuning and Optimization

- **Objective:** To fine-tune the chosen model's hyperparameters to achieve optimal performance and avoid overfitting.
- **Activities:**
 - Identify the key hyperparameters of your chosen model.
 - Apply hyperparameter tuning techniques, such as GridSearchCV or RandomizedSearchCV, to explore different parameter combinations.
 - Evaluate model performance on a validation set during tuning.
 - Select the hyperparameter values that result in the best performance.
 - Retrain the model with the optimized hyperparameters on the full training data.

Meet your educator:-

Hi I am Himanshu Arora, and I will be your instructor for your course. I have about 3 years of experience in Artificial Intelligence, Machine learning and various Data Science Courses. I have worked in Geeta University, Parul University, byteXL. I also work with different business organizations. I have a Bachelor's degree in CSE, Masters' in Data Science. When I am not teaching I enjoy Singing, playing Cricket & Basketball. I also love travel and photography.

About the Nimbus Platform:-

I already give main details in upper pages of this file, here I give you the future scope of byteXL Nimbus Platform in details.

Expansion of Course Offerings and Learning Paths:

- byteXL is likely to continue expanding its library of courses and learning paths, covering a wider range of in-demand skills and technologies. This will provide learners with more options to choose from and tailor their learning journey to their specific career goals.
- They might introduce new domains like cybersecurity, cloud computing, or data visualization, allowing learners to develop specialized skills in these areas.

Enhanced Interactive Learning Features:

- The platform could incorporate more interactive elements, such as gamification, simulations, and virtual labs, to enhance engagement and make learning more immersive.
- They might integrate features like real-time collaboration tools or code pairing functionalities to promote peer learning and teamwork.

Deeper Integration with Industry Tools and Technologies:

- byteXL might partner with leading technology companies to integrate their tools and platforms into the Nimbus environment. This would provide learners with hands-on experience using industry-standard software, making them better prepared for real-world job scenarios.
- They could establish connections with cloud providers like AWS or Google Cloud, allowing learners to deploy and manage projects in real cloud environments.

Personalized Learning Recommendations and Career Guidance:

- The platform could leverage data and AI to provide personalized learning recommendations to learners based on their skills, interests, and career aspirations. This would help learners make informed decisions about their learning paths and focus on the most relevant content.
- byteXL might integrate career guidance tools, such as resume builders or interview preparation resources, to support learners in their job search and career development.

Community Building and Collaboration:

- byteXL could foster a stronger sense of community among learners by creating forums, discussion boards, or online events where they can connect, share ideas, and collaborate on projects.
- They might introduce mentorship programs or peer-to-peer support networks to provide learners with guidance and encouragement.

Integration with Assessments and Certifications:

- byteXL could partner with industry organizations or certification bodies to offer recognized credentials upon completion of certain courses or learning paths. This would add value to the learning experience and enhance learners' employability.
- They might develop their own assessment tools or integrate with existing platforms to provide learners with feedback on their skills and progress.

Global Reach and Accessibility:

- byteXL could expand its reach to a wider global audience by offering courses and resources in multiple languages and adapting to different learning styles.
- They might introduce features to improve accessibility for learners with disabilities, ensuring inclusivity and equal learning opportunities.

