



QuestDice 
A quest for fun

Business Presentation

Welcome to QuestDice, Where Fun Knows No Boundaries! 🎲

At QuestDice, we are passionate about creating unforgettable gaming experiences that bring joy, laughter, and strategic thinking to families and friends around the world. With a dedication to quality, innovation, and community, we have established ourselves as an innovative creator of engaging board games that captivate players of all ages.

Our journey began with a simple belief: that the power of play can bring people together, raise creativity, and trigger memorable moments. Since our beginning, we have given our hearts and minds into developing an innovative game that cater to a wide range of interests and people.

What sets us apart is our relentless commitment to excellence in every aspect of game design. From meticulously crafted components to intuitive gameplay mechanics, we strive to exceed expectations and deliver an immersive gaming experience that delights both casual players and seasoned enthusiasts alike.

But our mission extends beyond just creating great games. We believe in fostering a vibrant community of gamers who share our passion for tabletop entertainment. Through engaging events, online forums, and interactive social media channels, we aim to connect players from all walks of life, sparking lively discussions, friendly competitions, and lasting friendships.

At QuestDice, we are more than just a game company — we are a gateway to adventure, laughter, and endless possibilities. Join us on this exciting journey as we continue to push the boundaries of creativity and innovation in the world of board gaming.

Experience a quest for fun. Experience QuestDice.

The project

Introduction

QuestDice was launched just one year ago with the sale of a single unique product, and its success was tremendous. The first board game created was sold to 198,723 customers.

In the initial phase for the second product launch, QuestDice sent direct advertising to a small portion of its previous customers. In the available dataset containing 15,589 responses, it was observed that out of the total, 8,477 customers made the purchase of the second product, while 7,112 did not. With the advertising cost per customer estimated at €3 and the profit per sell (excluding advertisements costs) at €8, the resulting profit amounted to €21,049 (calculated as: $8,477 * €8 - 15,589 * €3$). However, the company is considering adopting a more analytical approach, targeting customers with a higher likelihood of purchasing the product, with the aim of reducing costs.

Hopefully, with this approach, the revenue could be higher by sending more efficient advertising to the possible 198,723 previous customers who did buy the first product.

Objective

The main goal of this project is to create and apply predictive models to forecast customer purchasing patterns for the upcoming release of our company's second product, following the successful launch of the first product. Specifically, the project aims to predict the probability of customers purchasing the second product based on their survey responses and the data available for each customer. Through an analysis of customer data and survey feedback, the project aims to predict what are the customers more prone to buy the second product and increase the revenue generated.

Methodology

The project will employ a data-driven approach, leveraging techniques from machine learning and predictive analytics. Data was collected from customer surveys, which will capture various aspects of customer preferences, satisfaction, and purchasing behavior. Feature engineering will be performed to extract relevant variables from the survey data. Supervised learning algorithms, such as Nearest Neighbors, decision trees, and other methods, will be utilized to develop predictive models. The models will be trained on historical data and evaluated using metrics such as accuracy, precision, recall, and F1-score, with focus on F1-Score.

Expected Outcomes

Upon completion of the project, we anticipate the following outcomes:

- Development of predictive models to forecast customer purchasing behavior for the upcoming launch of the second product. By the usage of predictive analytics insights, the aim is to optimize resource allocation, resulting in enhanced return on investment (ROI) by reducing advertising costs. This involves targeting advertising efforts specifically towards customers with a higher probability of purchasing the second product, rather than promoting it to all customers indiscriminately.
- Identification of key factors influencing customer decisions to purchase a second product.

The data

The available data is composed of two datasets:

- **train.csv** - the training set (you can use this dataset as you wish in your notebook - just to train, you can make a train_test_split, you can apply Kfold or something similar...). It is composed of 15.589 observations.
- **test.csv** - the test set. Composed of 5.195 observations.

The available data comprises demographic and firmographic information, as well as responses to fourteen questions provided by customers regarding their perception of the first board game launched. These responses are graded on a scale from 0 (Completely Disagree) to 5 (Completely Agree). The target variable is 'Buy_Product,' where 0 denotes customers who did not purchase the second product, while 1 represents those who did.

In the test dataset, access to the target variable is unavailable. Therefore, the model constructed must predict and populate this variable.

Column	Description
Cust_ID	Unique identification of the Customer
Name	Customer's Name
Newsletter_Subscription	Indicates whether a customer has opted to subscribe to QuestDice's email newsletter.
Year_Birth	Customer's Year of Birth
FidelityPoints	Customer's fidelity points obtained through various interactions and engagements with the company, including purchases of board games, participation in promotional activities, and engagement with digital platforms such as the company's website or social media channels.
Preferred_Game_Genre	Customer's gaming preference (Strategy, Party, Thematic)
Membership	Costumer's membership type (Premium or Non-premium)
S1_GameplayExperience	Answer to question: "I found the gameplay engaging and enjoyable."
S2_GameMechanics	Answer to question: "The game rules and mechanics are complex and innovative"
S3_GameBalance	Answer to question: "The game is fair and balanced for all players."
S4_GameLength	Answer to question: "The duration of gameplay sessions is appropriate."
S5_MaterialQuality	Answer to question: "The game pieces, cards, board, etc., are of high quality."
S6_ValueForPrice	Answer to question: "I perceive the value of the game to be worth the price paid."
S7_Artwork_Design	Answer to question: "The visual appeal and thematic artwork quality are high."
S8_ThemeIntegration	Answer to question: "The theme is well integrated with gameplay mechanics."
S9_PlayerIteration	Answer to question: "The game incentivizes the iteration among players."
S10_LevelAdjust	Answer to question: "The game tailors well to different age groups or skill levels."
S11_RuleClarity	Answer to question: "The game rules are clear and understandable."
S12_StrategicThinking	Answer to Question: "The game encourages strategic thinking and decision-making."
S13_ReplayValue	Answer to question: "I am likely to replay the game."
S14_Recommendation	Answer to question: "I am likely to recommend this game to others."
Buy_product	The target (0 – Did not buy the second product; 1 – Buy the second product)

Deliverables and Evaluation

1. A **Jupiter notebook** with all the needed code implemented to obtain the results presented in the report and to obtain the results explored in the report.

The file naming format should be "DSML202324_Predictive_GroupXX_Notebook.ipynb", where "GroupXX" should be your group number.

2. A **report** that describes the analytical processes and the conclusions obtained. A project that focuses only on the techniques and methodologies approached during the practical classes will have at most 15 values. The remaining 5 values are possible to achieve if contributions based on self-study and creativity are applied, and clearly explained in the report.
 - a) It should contain the following structure (text in figures and tables do not count as words):

CHAPTER	MAXIMUM WORDS	MAXIMUM POINTS
Kaggle Performance	NA	5 (25%)
Abstract	250	0.5 (2.5%)
1. Data Exploration		1 (5%)
2. Preprocessing		2 (10%)
3 Modelling and Fine-Tuning	3000	2 (10%)
4 Performance Assessment		1 (5%)
5 Conclusion		1.5 (7.5%)
References	NA	2 (10%)
Report Quality and StoryTelling	NA	
Creativity and Other Self-Study (Optional)		
Annex1 – MIC feature selection	400	1 (5%)
Annex 2 – Other predictive Models (excluding Ensembles)	1000	1.5 (7.5%)
Annex 3 – Ensemble Models	1000	1.5 (7.5%)
Annex 4 – Creativity and other Self-Study	500	1 (5%)

- b) The font formatting should respect the following conventions:
 - Heading 1: Arial, Size 12 pt, in bold
 - Heading 2 (if needed): Arial, Size 11 pt, in bold and italic
 - Text: Arial, Size 11 pt, line space of 1.5 points.
 - Margins: The default ones in word (Top, Bottom, Left and Right as 1").
- c) The reports that do not follow the specified conditions will suffer penalization on the grade.
- d) The file naming format should be "DSML202324_Predictive_GroupXX_Report.pdf", where "GroupXX" should be your group number.

- e) All chapters and moments of evaluation are graded through a comparison of the work provided by the different groups.

Notes

- The deadline for the project is the 15th of June 2024. You may submit your project (notebook and report) up to three days after the deadline, until the 18th of June 2024. However, for each day of delay, a deduction of one point (out of 20) will be applied to the final evaluation score.
- We will evaluate all the topics mentioned based on the report - a well-structured and succinct report will have a big weight on the evaluation.
- The jupyter notebook will be analyzed only if some doubt arises during the report evaluation. If some steps were done in the Jupyter notebook but not described in the report, we will not evaluate those. As an example, imagine you check the outliers, and at the end of your project, you decide to keep them. In the report, you should mention how you check if you had outliers, what steps were taken to remove them and why you decide to keep them in the end, among other insights that can be relevant.
- The jupyter notebook should be delivered with all the cells already run and the outputs visible.
- The report and the code will pass through a process of plagiarism checking.
- Theoretical context about algorithms / techniques applied should only be provided when those approaches were not given during the practical classes.
- The report should clearly refer (on a cover or on the first page if no cover is included) to the group number, the students' names and the students' numbers.
- In Kaggle you should assign one and only one submission as your final submission, that will correspond to the notebook delivered and to the model you refer as your final model in the report (more instructions on how to select only one submission on moodle - "How to submit and select the final submission on Kaggle").
- One submission (report and notebook) per group is enough.

For more information, please read the Kaggle competition rules carefully.

Chapter Description

This bullet-list provides some details about each aspect:

- **Kaggle performance (*not a chapter*):** The performance obtained on Kaggle, on the submission selected (F1 Score).
- **Abstract:** A small summary of your work. The abstract should give an overview of your work: What is the context? What is your main hypothesis? What did you do? What were your main results and what main conclusions did you draw from them.
- **Data Exploration:** Describe the data available and extract meaningful insights that may be helpful in addressing the problem at hand.
- **Preprocessing:** This stage includes all the steps from raw data into data ready for modelling: data cleaning, transformation, and reduction. It also entails business-related transformations of the input features, creating new features if feasible, and accompanying explanations, and the feature selection stage. Here, we evaluated the quality of your implementation, its justifications, and the insights extracted from this stage.
- **Modelling and Fine-Tuning:** the implementation of different predictive algorithms and the process of fine-tuning those models. The application of additional models not given during classes are optional and considered as points in “Other predictive models” or “Ensemble Models”, depending on the type of algorithms implemented.
- **Performance Assessment:** The comparison of different models and their performance.
- **Conclusion:** Summarizes the key supporting ideas you discussed throughout the work. This conclusion should also focus on the primary goal of the project, discussing on the impact that the application of the final model will have on the financial return.
- **Report-quality and Storytelling (*not a chapter*):** Each report should follow the provided report structure and describe the steps and main insights along the process. Clarity, synthesis, objectiveness, and business-contextualization are very welcome. Your decisions and steps must be reasonably justified by the previous findings (when this is possible and feasible), your hypothesis and findings must be related to the problem’s business-context, etc.

ANNEXES: CREATIVITY AND OTHER SELF-STUDY TOPICS(OPTIONAL)

If other techniques not covered during practical classes are applied, a theoretical explanation of the algorithm/technique should be provided. In these topics, it is expected that students explore other concepts and techniques not covered during the theoretical and/or practical classes, not only in the theoretical context and implementation but also in the analysis of the outputs and possible comparison of those with the default ones applied in the remaining chapters.

- **MIC feature selection:** The application of Mutual Information Coefficient as an additional approach to select the most important features to include in the model. A theoretical explanation of this technique should be provided in Annex 1.
- **Other predictive models (excluding ensemble techniques):** A theoretical explanation of the algorithm(s) should be provided in Annex 2. Involves the depth and the quality of the comparative analysis provided by the different algorithms, the theoretical explanation of the algorithm(s) itself and the justification of the chosen hyperparameters. These models should be exclusively sourced from the sklearn library.
- **Ensemble Techniques:** A theoretical explanation of the ensemble algorithm(s) should be provided in Annex 3. Involves the depth and the quality of the comparative analysis provided by the different algorithms, the theoretical explanation of the algorithm(s) itself and the justification of the chosen hyperparameters. These models should be exclusively sourced from the sklearn library.
- **Creativity and other Self-Study:** This topic includes not only the application of different techniques not defined on the previous topics (with a theoretical explanation of the same if feasible) but also aspects of creativity, such as the quality of visualizations, plots and others.

All topics are evaluated through a comparison of the work provided by the different groups.