M1 Requirements Document: In-Game Purchases Prediction

Group 3 - Team members

- Vaishnavi Kokadwar
- Mahender Reddy Pokala
- Kunal Mody
- Ulka Khobragade
- Ultra Partihuttakorn

Project Leadership

Project Lead - Vaishnavi Kokadwar

The Project Lead provides strategic direction and ensures alignment with project objectives. Key responsibilities include:

- Setting project vision and objectives
- Making critical technical decisions
- Managing stakeholder communications
- Ensuring deliverable quality
- Coordinating with external stakeholders

Scrum Master - Ulka Khobragade

The Scrum Master facilitates the development process and removes impediments. Core duties include:

- Facilitating daily stand-ups and sprint meetings
- Maintaining the project board and backlog
- Removing team blockers
- Ensuring adherence to agile practices
- Supporting continuous improvement

Technical Team Composition

Data Collection Lead - Mahender Reddy Pokala

- Leads data acquisition and preprocessing
- Implements data cleaning pipelines
- Ensures data quality standards
- Oversees feature extraction processes

Model Development Lead - Kunal Mody

- Designs machine learning model architecture
- Conducts model selection and tuning
- Evaluates model performance
- Implements optimization strategies

Feature Engineering Specialist - Ulka Khobragade

- Develops feature transformations
- Conducts feature importance analysis
- Creates feature selection pipeline
- Implements dimensional analysis

Model Evaluation Lead - Ultra Partihuttakorn

- Implements evaluation metrics
- Conducts validation testing
- Develops interpretability analysis
- Ensures model reliability

Data Pipeline Lead - Vaishnavi Kokadwar

- Ensures efficient data processing
- Implements deployment workflow
- Manages model versioning
- Maintains technical documentation

Case Description

The Online Gaming Behavior Prediction System aims to develop a machine-learning model to predict player engagement levels and optimize game performance. The project follows an agile methodology, with two-week sprints, to deliver a production-ready model within four weeks.

Machine Learning Objectives

1. **Predictive Accuracy**
    - Achieve 85% accuracy in engagement level prediction
    - Implement cross-validation and stratification
    - Generate probability scores for predictions
2. **Feature Engineering**
    - Process demographic and behavioral features
    - Transform temporal gaming metrics
    - Identify key engagement indicators
3. **Model Development**
    - Implement classification models (Random Forest, XGBoost)
    - Compare model performance
    - Optimize for multi-class prediction
4. **Model Interpretation**
    - Analyze feature importance for engagement
    - Create player segment profiles

Agile Implementation Framework

- Weekly reviews and planning
- Continuous backlog refinement
- Sprint retrospectives

Business Success Metrics

1. Player retention improvement
2. Revenue impact prediction
3. Model inference speed

Four-Week Schedule with Final Presentation

Week 1: Data Preparation

- Data collection and initial cleaning
- Basic preprocessing pipeline
- Initial feature extraction
- Deliverable: Cleaned dataset and preprocessing pipeline

Week 2: Feature Engineering and Initial Modeling

- Core feature engineering
- Feature selection
- Initial model implementation
- Deliverable: Feature engineering pipeline and baseline model

Week 3: Model Development and Optimization

- Model refinement
- Basic hyperparameter tuning
- Initial validation
- Deliverable: Optimized model with documentation

Week 4: Final Testing and Presentation

- Complete final model validation
- Prepare comprehensive performance metrics
- Create presentation materials
- Rehearse and deliver the final presentation
- Deliverable: Final presentation and complete project documentation

Dataset

Source:
https://www.kaggle.com/datasets/rabieelkharoua/predict-online-gaming-behavior-dataset

Features:

i. **PlayerID**: Unique identifier for each player.
ii. **Age**: Age of the player.
iii. **Gender**: Gender of the player.
iv. **Location**: Geographic location of the player.
v. **GameGenre**: Genre of the game the player is engaged in.
vi. **PlayTimeHours**: Average hours spent playing per session.
vii. **InGamePurchases**: Indicates whether the player makes in-game purchases (0 = No, 1 = Yes).
viii. **GameDifficulty**: The difficulty level of the game.
ix. **SessionsPerWeek**: Number of gaming sessions per week.
x. **AvgSessionDurationMinutes**: Average duration of each gaming session in minutes.
xi. **PlayerLevel**: Current level of the player in the game.
xii. **AchievementsUnlocked**: Number of achievements unlocked by the player.
xiii. **EngagementLevel**: Categorized engagement level reflecting player retention ('High', 'Medium', 'Low')

Target Variable:

xiv. **EngagementLevel**: Indicates the level of player engagement categorized as 'High', 'Medium', or 'Low'.