1. Introduction and topic and/or problem statement – A short introduction and summary of the goals of the project
2. Data sources that will be used – A reference to any datasets utilized in the project
3. List of high-level methods, techniques and/or technologies that you are considering using.
4. Products to be delivered – what are the primary deliverables for the project? This is what we will be grading

**Introduction and Topic/Problem Statement**

The goal of this project is to develop a sentiment analysis model that can effectively classify the sentiment expressed in user reviews of video games. By analyzing the sentiment of game reviews, we can gain valuable insights into the aspects of games that users enjoy or dislike, which can inform game development and marketing decisions.

This project will focus on collecting a dataset of video game reviews from various online sources, preprocessing the text data, and training machine learning models to detect the sentiment expressed in the reviews. We will explore both traditional sentiment analysis techniques as well as more advanced transformer-based language models to identify the most effective approach for this domain.

Additionally, we will investigate aspect-based sentiment analysis, where the model can identify the sentiment expressed towards specific game features, such as graphics, gameplay, story, and performance. This level of granularity will provide deeper insights into the users' experiences.

**Data Sources**

The primary data sources for this project will be:

1. Professional video game review websites (e.g., IGN, GameSpot, Metacritic, kagg;e)

2. User review platforms (e.g., Steam, Google Play Store, App Store)

We will collect a representative dataset of reviews covering a diverse range of game genres, platforms, and release timelines. The dataset will be balanced in terms of positive, negative, and neutral sentiments to ensure the model can effectively learn to distinguish between them.

**Methods and Techniques**

The key methods and techniques we plan to use for this project include:

1. **Text Preprocessing**: Cleaning and preparing the review text data, including handling gaming-specific terminology, abbreviations, and user-generated content challenges.

2. **Feature Engineering**: Extracting relevant features from the review text, such as sentiment-bearing words and phrases, to be used as input to the machine learning models.

3. **Sentiment Analysis Models**: Experimenting with both traditional machine learning algorithms (e.g., logistic regression, decision trees, support vector machines) and transformer-based language models (e.g., BERT, RoBERTa) for sentiment classification.

4. **Aspect-based Sentiment Analysis**: Developing models to detect the sentiment expressed towards specific game aspects, such as graphics, gameplay, story, and performance.

5. **Model Evaluation**: Assessing the performance of the sentiment analysis models using appropriate metrics, such as accuracy, F1-score, and precision-recall curves.

6. **Visualization and Insights**: Analyzing the sentiment trends and patterns in the data and presenting the findings through informative visualizations and summaries.

**Products to be Delivered**

The primary deliverables for this project will include:

1. A Jupyter notebook that demonstrates the entire workflow, from data collection and preprocessing to model training, evaluation, and deployment.

2. A comprehensive report that covers the following:

- Problem statement and motivation

- Data collection and preprocessing

- Model architecture and training details

- Evaluation results and comparative analysis

- Insights and findings, including aspect-based sentiment analysis

- Potential real-world applications and future work