January Deliverable

1 Progress Made

Overall, progress made has been better than expected with 2 large sections completed and the most important section (The Sentiment Analysis Model) completed as a first draft. I will go back and refine the sections once I have received formal feedback but I am happy with where they are at the moment. Completing so much of the writing is a big step and should hopefully allow me more freedom to build on what I have actually developed closer to the deadline as I won't have to worry about writing so much.

The first proper stage completed was the literature review which was completed around 6 weeks into the first semester. I am happy with the research I was able to find surrounding the topic as I was unsure if such a niche topic would result in some problems in that area. Thankfully there are plenty of similar studies and additional research for the wider field of sentiment analysis which gave me much to write about. It also taught me a lot of useful information that made developing the sentiment analysis model much easier.

The development of the sentiment analysis model took a bit longer than initially expected, but I had not planned to do a write up for it this early so it has balanced out. I was able to fully research, implement and evaluate the sentiment analysis model achieving very good accuracy scores. This success makes me hopeful the approach I am taking can work and result in a prediction model that performs better than those I have seen in my literature review.

As well as this I have written an introduction and a explainer chapter for readers unfamiliar with the FPL game. This should hopefully make the report more readable and enjoyable for people new to the genre. It should also make the report a lot more clear in terms of any jargon or concepts that are not initially obvious.

2 Problems encountered

By far the biggest problem I encountered was the difficulty of obtaining twitter data. As this topic is so niche, it is hard to find good quality public datasets that are relevant and pre-labelled accurately. I was initially hoping to use the

twitter API to retrieve real-time tweets and create an autonomous system, but I found the price to be $100 a month which is not something I can afford.

I then also realised that this meant I could not use this current season's FPL data as the only twitter data I can find is for previous seasons, so I would not be able to match sentiment analysis scores to this current season. This means it will be harder to go back and find historical player data (like goals scored, minutes played, red cards etc.) that matches the timestamps of my tweets to properly evaluate the final model. However, this should be doable given that data is needed for all things football related (not just FPL) and is very accessible.

I also had the problem of not being able to find hand-labelled data meaning some of the training labels are inaccurate. I stuck with this approach as a tradeoff between either that or having to manually label data which would have required lots of time and a labelling criteria. It would also have reduced the dataset size I would have been able to use, resulting in worse model performance. This is mentioned in the relevant section of my report.

3 Changes of Direction

Given the problems encountered with obtaining twitter data, I have changed the main goal of the project from creating an autonomous system that uses real-time data, to investigating the performance gains that can be achieved when simulating the models I create on the historical data I can obtain. This is not ideal from a user experience point of view as no product is being created, but it will still contribute to the existing research surrounding this topic and the main parts of the system will still be the same. The system will be developed keeping in mind potential to integrate the twitter API and create an autonomous system, if sufficient funding can be secured.

This also means there is no need for an SQL database to keep current season data, as there are no tweets to go with them. Similarly there is no need for integration with the FPL API to automatically manage a team, and there is no need to scrape data from existing AI platforms as their current data will not be used.

To make up for this decrease in workload, I have decided to create my own expected points prediction model that uses the sentiment analysis model's output as an input parameter. I was unsure how to use the output of the sentiment analysis for picking an FPL squad, initially considering just taking an average of the output and adding or subtracting it to an expected points prediction from a third party. However, instead of blindly trialing different values and weights for this method, using my own model should correctly weight the sentiment analysis score in terms of an overall prediction, meaning it will not be too heavily relied on, or ignored too much. It should also allow me to better compare my results to other existing models by using the exact same input parameters and then adding sentiment analysis to see what difference it makes.

4 Updated Work Plan

	1/25			2/25			3/25			4/25		
	7 12 19	26	2 9 16	23	2 9	16	23 30	6	13 20	27		

fyp

Exams
    Exam Revision

Expected Points (xP) model
    Research existing models
    Obtain suitable dataset
    Pre-process data
    Train model
    Optimise hyperparameters
    Evaluate model

Team Selection Algorithm
    Research Existing Algorithms
    Select best algorithm
    Define FPL constraints
    Implement algorithm
    Test algorithm

Combine Everything to simulate pe...
    Find matching twitter data and playe...
    Use output of SA model for xP model
    Use Team Selection Algorithm to pick...
    Record Performance of chosen team

Report Write up
    General Writing and Refining
    Final Writing of Results and Discussi...
    Report Submission