**Capstone Project Proposal**

**Company**: Covers.com (website giving information about online sports gambling)

**Business Goals**: Investigate whether or not we can create a picker service. We will charge a monthly fee for this service if it is found to be useful. For this proof of concept project, we will focus on the National Basketball Association (NBA).

**Notes**: The picker service will use a machine learning algorithm to try and pick the game’s winner. The project will have two teams. The machine learning team and data gathering team.

The machine learning team will ……..

The data gathering team will find data sources available on the web News stories will be broken down into word counts and frequencies, and game data will be stored in an appropriate format.

**API(s)**: NBA-API (https://api-nba-v1.p.rapidapi.com/)

ESPN NBA RSS Feed – (https://www.espn.com/espn/rss/nba/news)

Perhaps others if time permits

**Pipeline tools**:

API(s)

Java/Python/Scala **code** as consumer(s) of API and producer(s) to Kafka

github – to store code

Jenkins – to <propagate code to the server ASAP>

Kafka

Buffer data until spark is ready

Spark Structured Streaming

Receive data from Kafka

Make necessary transformations on received data

Word Counts

Statistical results

Hive database as storage for transformed data which will be used by machine learning team

Machine Learning Team

works magic and creates best pick(s) of the day

Visualization tools

Machine Learning Team’s results incorporated into our new pay website