**Capstone Two - Project Proposal**

**Movie recommender system**

**Problem Statement Worksheet (Hypothesis Formation)**

|  |
| --- |
| Provide a movie recommendation to a user based on other movies that they have reviewed along with the ability to specify the genre, if needed |

**Context**

|  |
| --- |
| It’s a common use case where a user might be looking for a movie recommendation based on their taste without browsing through the entire library of movies. Therefore, the need is to build a recommender system which based on past review history could recommend a movie to the user.  Given that we have a pre-defined movie dataset, first the user could be asked to review some of those movies and then the recommendation can be provided. |

**Criteria for success**

|  |
| --- |
| 1. To be able to accurately recommend a movie to the user. 2. The accuracy can be tested against the review dataset available. 3. Feedback option can be provided to the user to react to the provided recommendation when put to real test. |

**Scope of solution space**

|  |
| --- |
| Recommend a movie out of the available dataset |

**Constraints within solution space**

|  |
| --- |
| 1. Movie recommendation can be provided out of the available movie dataset only. 2. Recommendation will be based on the reviews available in the dataset. |

**Stakeholders to provide key insight**

|  |
| --- |
| 1. Sunischal Dev – Project Mentor 2. Family and friends – To act as the users of the project to test the accuracy in real life |

**Key data sources**

|  |
| --- |
| [The Movies Dataset | Kaggle](https://www.kaggle.com/rounakbanik/the-movies-dataset)   1. The dataset is available as .csv files hence easy to consume 2. Size of the dataset is around ~900 MB 3. Attributes include:  * Keywords * Credits (like Department, Gender, Job) * Metadata like Title, Budget, Genre, Homepage, Original Language, Overview text * Ratings from over 700 users on 9,000 movies * IMDB and TMBD IDs |