AML 2304 – Natural Language Processing

Movie Sentiment Analysis

(Final Submission)

December 6, 2023

Submitted to :
Prof. Bhavik Gandhi

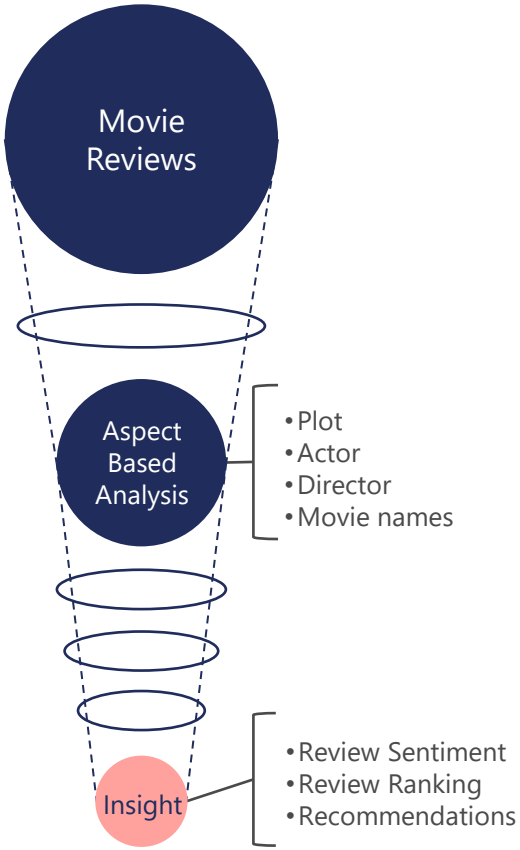
Submitted by Group 3 :
Abhishek Natani
Auradee Castro
Bhumika Rajendra Babu
Miraj Sinya
Olivia Deguit
Rochan Mehta
Roger Mais
Varun Sharma

Lambton College

CONTENT

Case Background

- Movie Reviews have always been a reference point for the audience to decide weather or not to watch movies but from a production standpoint it has not be utilized to its full extent.
- This application will allow the producers in the mass media and entertainment industry to not only understand the sentiment of the audience but also the reasons behind those sentiments.



Objectives:

- To make sense of the sentiment from the movie reviews and get the intent behind those sentiments giving deeper insights into audience opinions, preferences, and pain points leveraging Natural Language Processing Models.
- To get insights on trends in the audience's sentiment towards specific genre over the years, actors, directors which will aid them in making informed decisions.

Project Breakdown

Data Acquisition

Application Feature Development

Stage 1 | Stage 2 | Stage 3

Application Deployment

Activities

- Sourcing relevant data to address the identified business problem
- Cross-validating dataset
- Dataset size: **500k rows**
- **Data Sources used:**

kaggle

IMDb

- Training & Evaluating Sentiment Analysis Models

- Generating Sentiment labelled dataset using the tuned Sentiment Analysis Model

- Leverage the labelled dataset & trained model to build application features

- Design the application UI integrating the models and application features.
- Deploy the application
- Tools used:

Streamlit | python

- Sentiment Analysis Model building and Evaluation as the initial step for application feature development
- 4 features offered by the application to the users:

Feature1: Sentiment extraction from movie reviews
  - Sentiment Analysis using Trained Sentiment Analysis Model

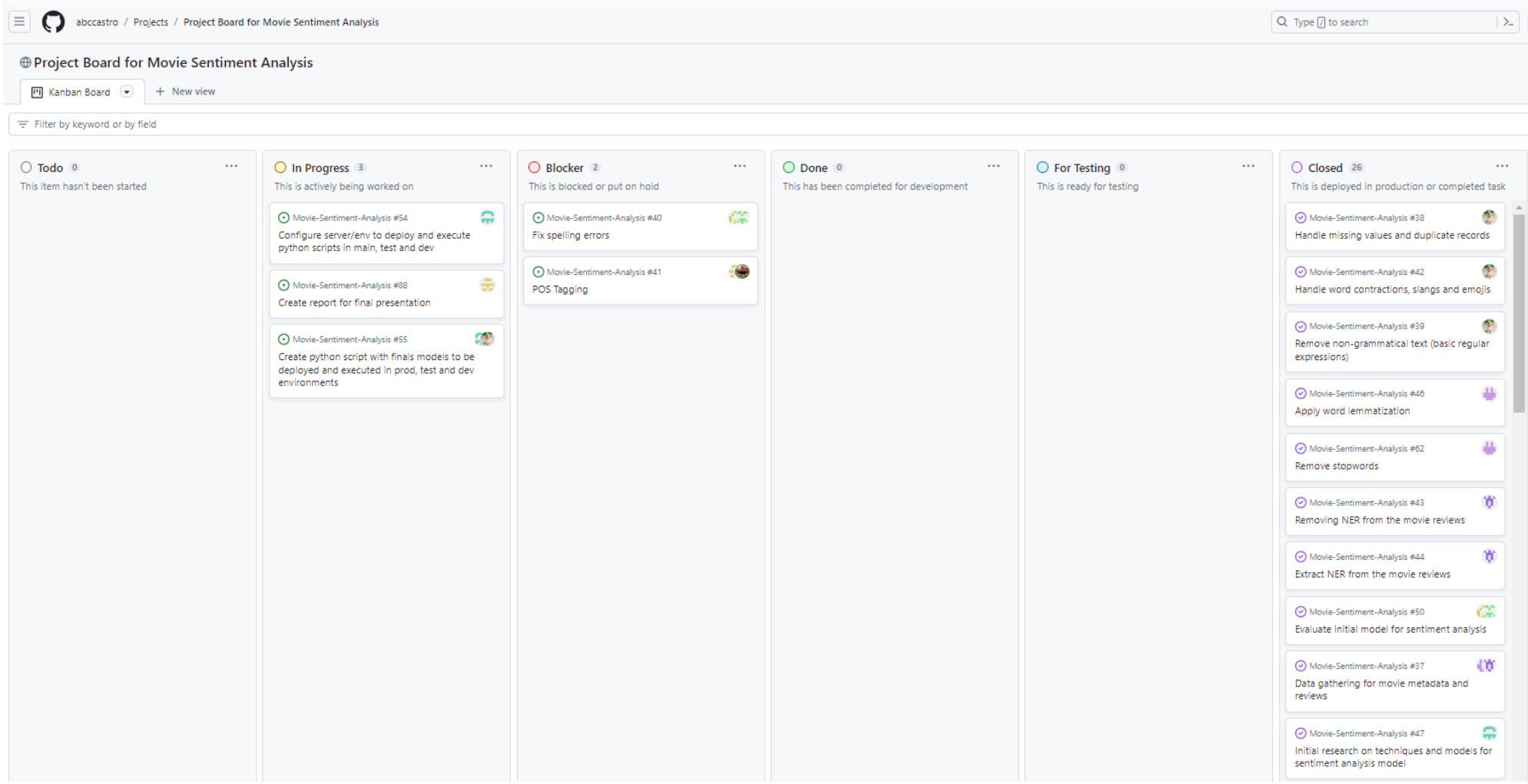
Feature 2 : Aspect based Sentiment Analysis
  - Sentiment Analysis based on user specified aspect(query)

Feature 3 : Movie based query
  - Sentiment Analysis for movies specified by the user

Feature 4 : Movie Recommendation
  - Movie recommendations based on common sentiment, rating, genre, actors, directors

Project Board Walkthrough

Using Github's Project to create Kanban Board for model development



* Project Board for Movie Sentiment Analysis: https://github.com/users/abccastro/projects/1

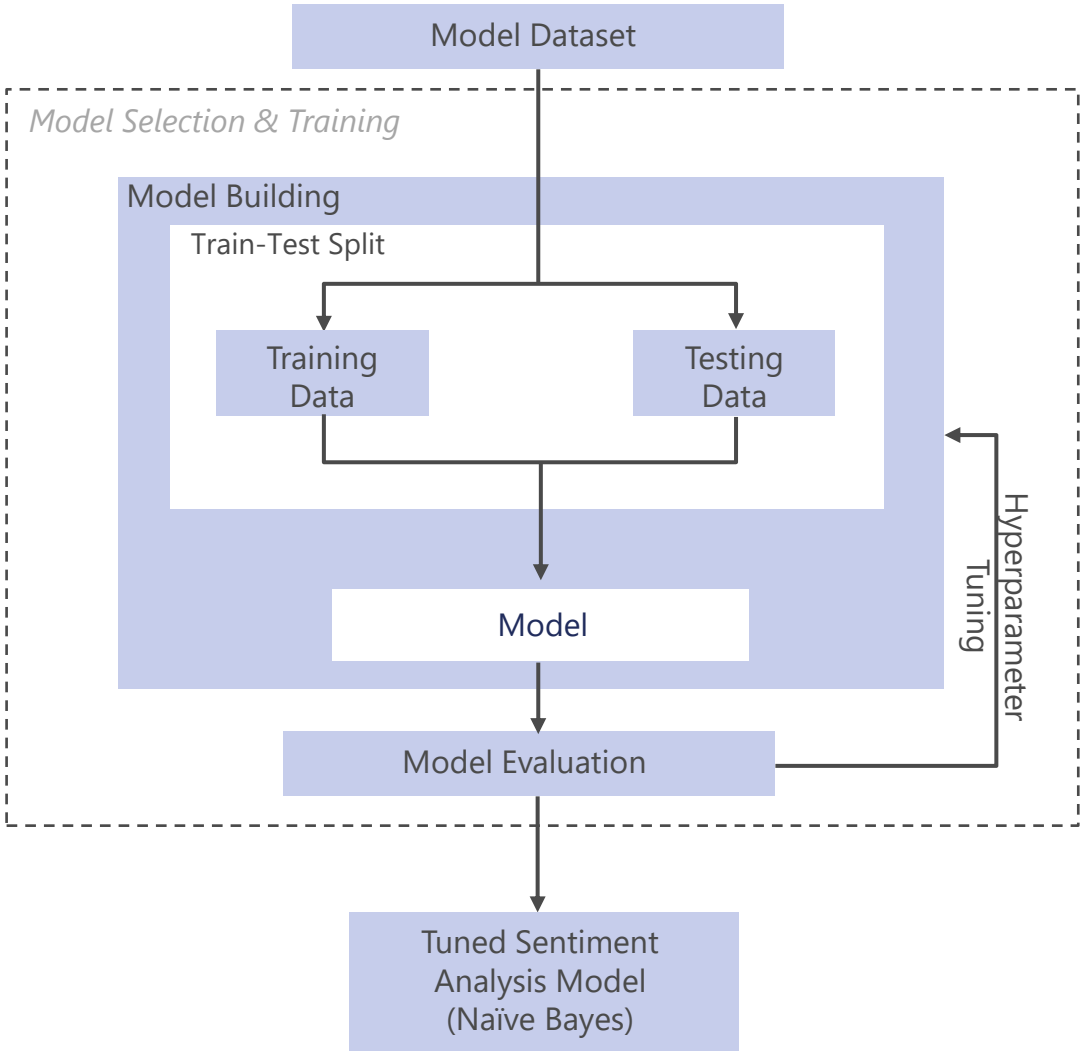
Team's Best Practices

- **Feature Breakdown:** Dividing features into smaller tasks for better prioritization and more manageable components
- **Distributed Responsibilities:** Assigning tasks to individual team members to ensure clear responsibility and accountability
- **Project Progress Checkpoint:** Conducting regular team meetings, via MS Teams or in person, for updates, blockers and planning next task
- **Model Development Workflow:** Leveraging * GitHub as a model development repository, adhering to industry standards with distinct branches for production, testing, development, and tasks, and conducting code reviews before merging

* Movie Sentiment Analysis repository: https://github.com/abccastro/Movie-Sentiment-Analysis

Sentiment Analysis Model : Model Building & Evaluation

Data Flow



Details

Model Creation and evaluation

- 3 Models trained using the labelled model dataset:
  - Multinomial Naïve Bayes
  - Long Short-Term Memory (LSTM)
  - Gated Recurrent Unit(GRU)
- Model Evaluation

Model	Accuracy
Naïve Bayes	**71%**
LSTM	59%
GRU	58%

  - Naïve Bayes performed the best with the accuracy of 71%

Hyperparameter Tuning

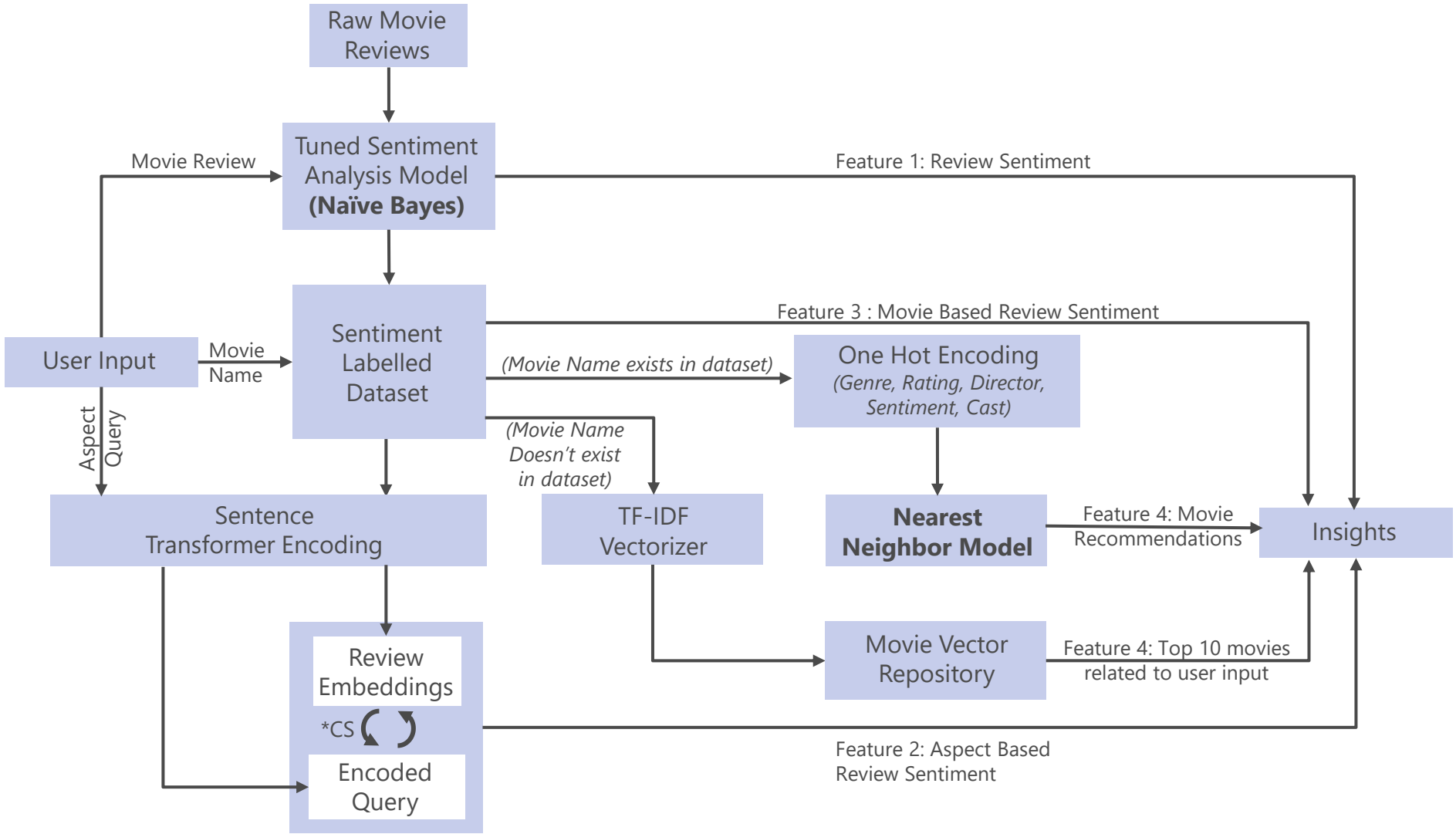
- GridSearchCV for hyperparameter tuning
- Parameter: Alpha

Note: Minor change when adjusting alpha; higher alpha values improve accuracy, while lower alpha values improve precision. Alpha 1 has an optimal balance across accuracy, precision, recall, and F1-score

* ROC (Receiver Operating Characteristic) and AUC (Area Under the ROC Curve)

Application Features Overview

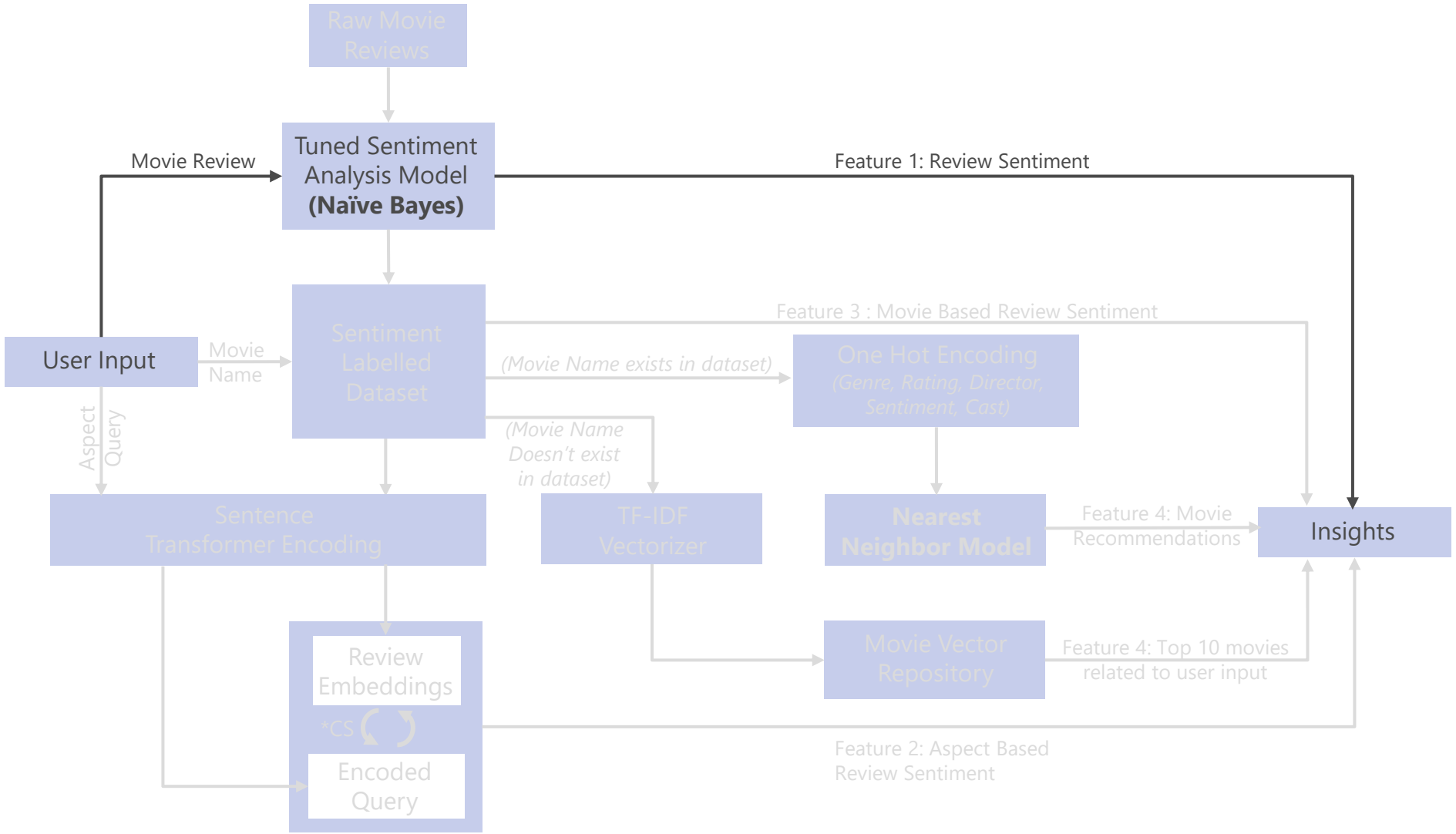
Implemented Naïve Bayes & KNN model to deliver Final Application Features



* CS : Cosine Similarity

Feature 1 : Review Sentiment

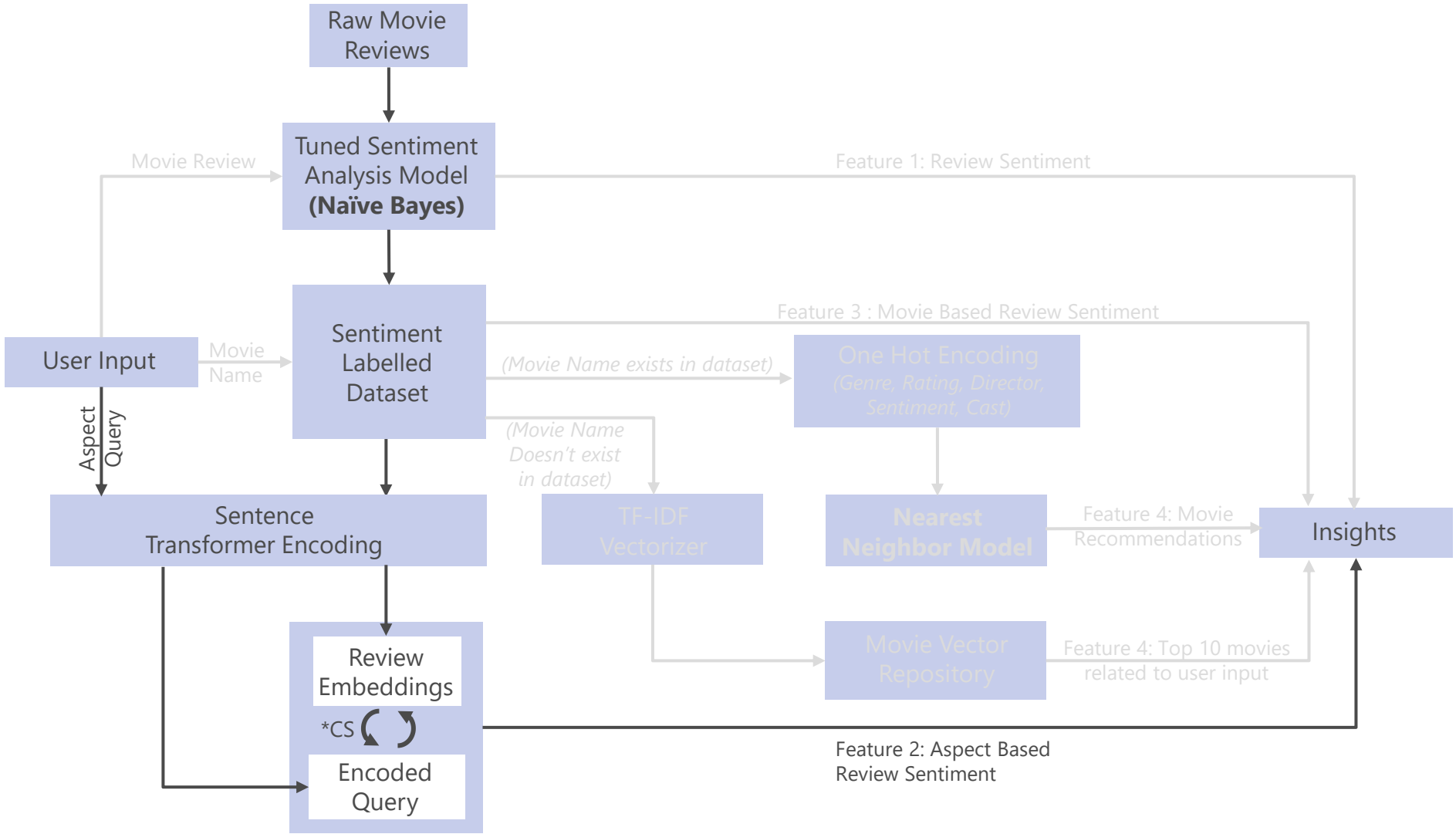
Implemented Naïve Bayes to give sentiments on reviews provided by the user



* CS : Cosine Similarity

Feature 2 : Aspect Based Review Sentiment

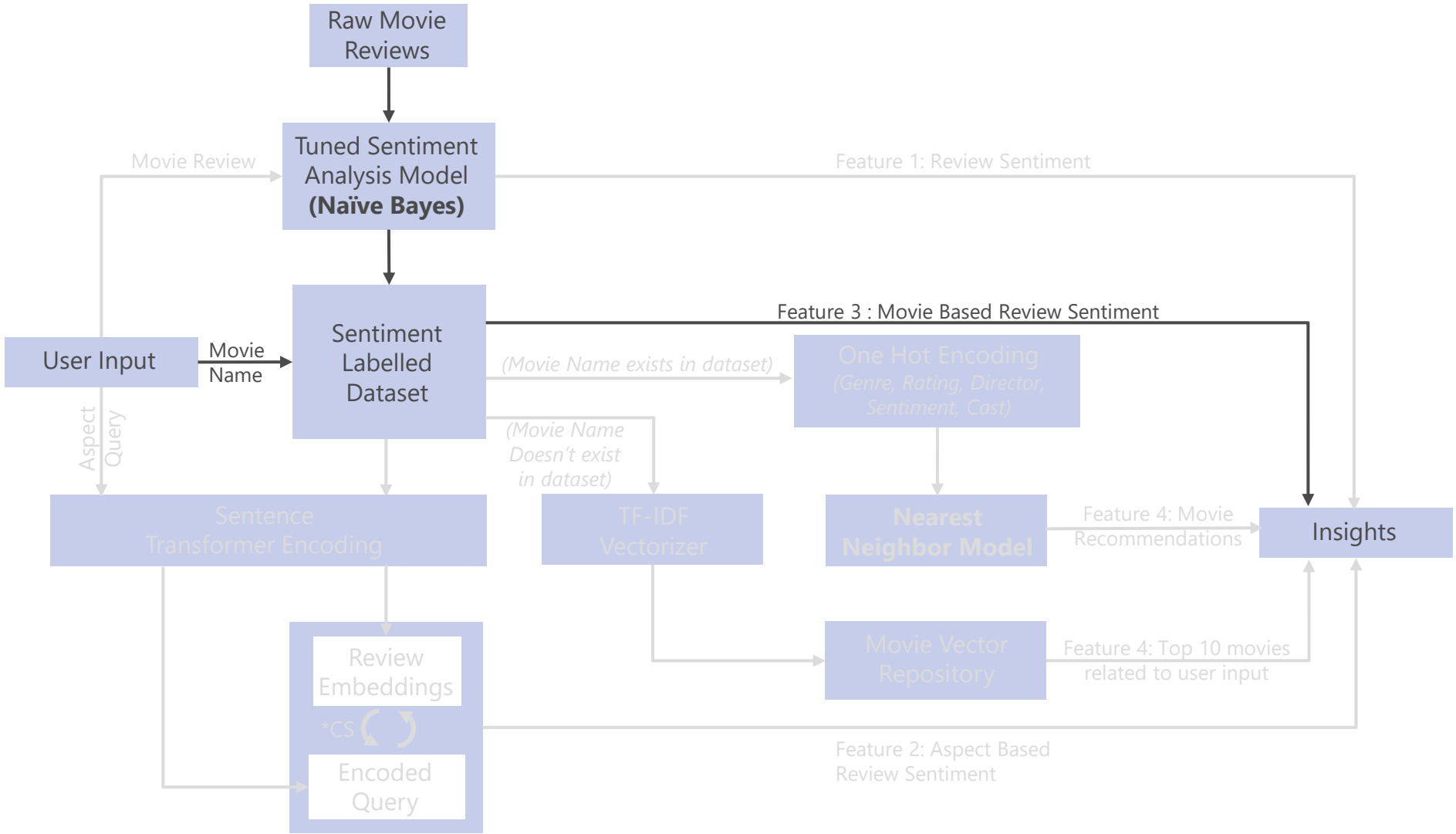
Implemented Naïve Bayes & Sentence Transformer to deliver Feature 2



*CS : Cosine Similarity

Feature 3 : Movie Based Review Sentiment

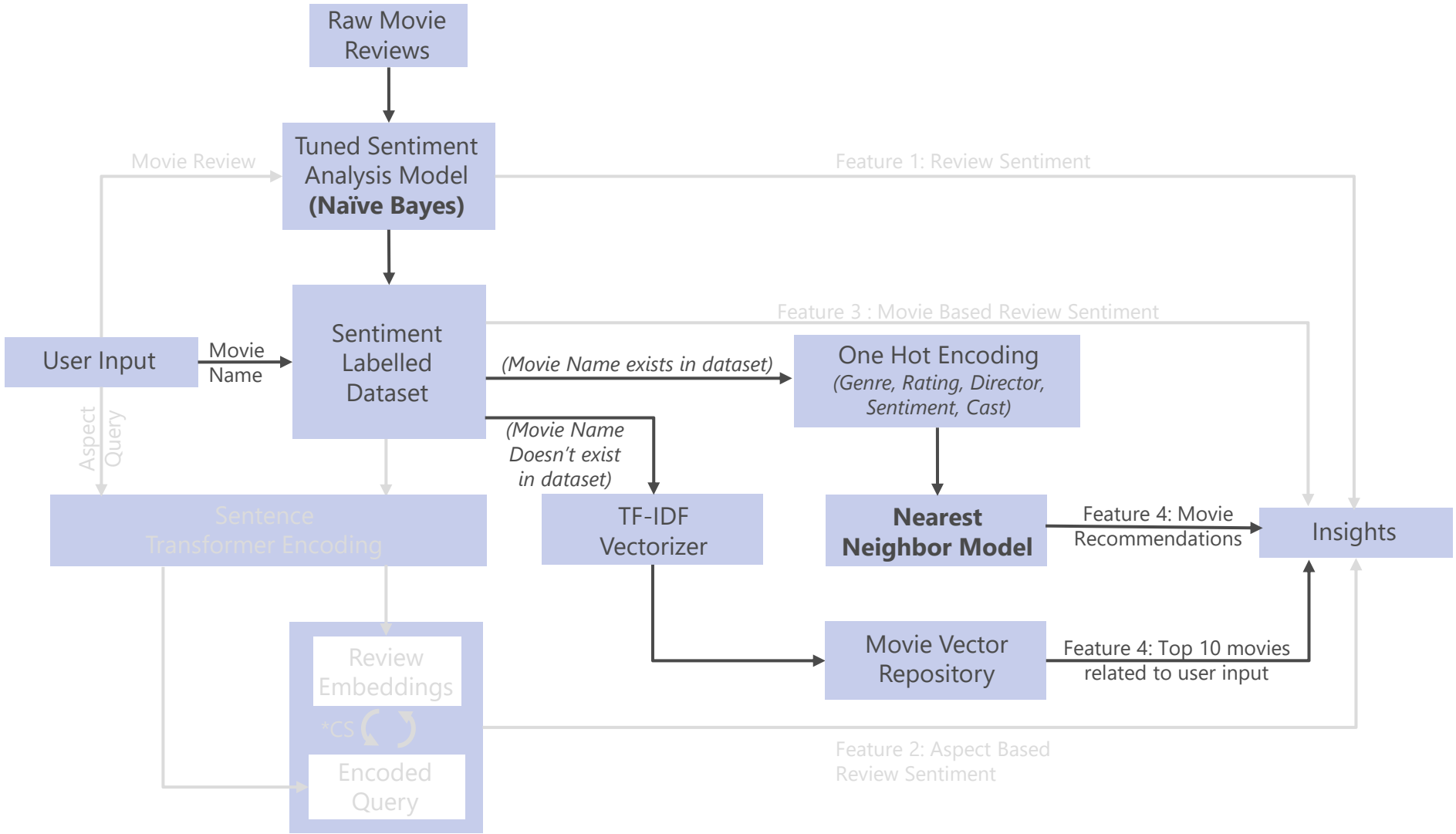
Filtering Sentiment Labelled Dataset based on user specified Movie Titles



*CS : Cosine Similarity

Feature 4 : Movie Recommendations

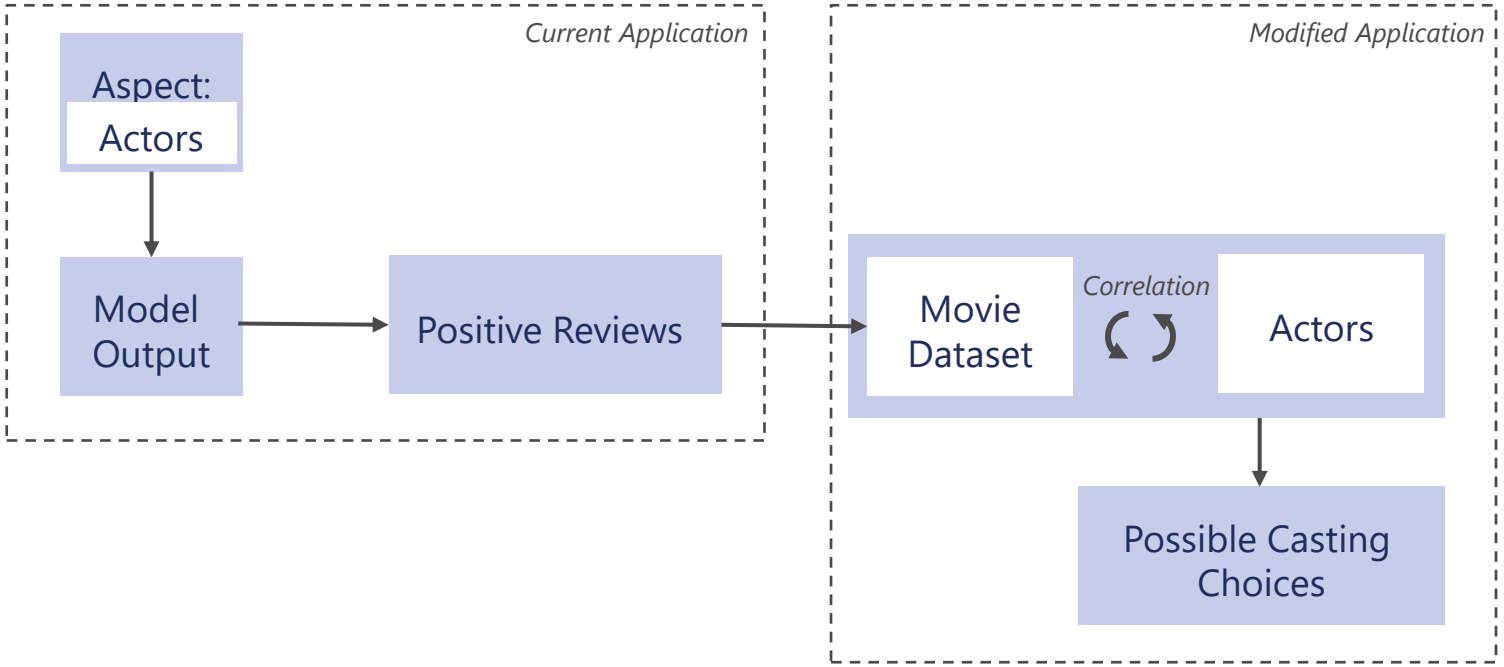
Implemented KNN model to deliver Feature 4



* CS : Cosine Similarity

Future Improvements

- Add the movie review sentiments generated from Feature 1 to the application's sentiment labelled dataset.
- Further filtering out the results from features 2,3 and 4 to get actors, directors, writers' correlation to the positive sentiment based on the existing model's output.




Lambton
College

References

- Banik, R. (n.d.). The Movies Dataset. Kaggle. https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset
- Zhang, X. (n.d.). IMDb Vision and NLP Dataset. Kaggle. https://www.kaggle.com/datasets/raynardj/imdb-vision-and-nlp/data