

## Markell2 Report

# PART1

### Step 6

#### Top 20 words for positive reviews

| No | Review | Clean Review | No Stopwords | Lemmatized |
|----|--------|--------------|--------------|------------|
| 1  | the    | the          | film         | film       |
| 2  | and    | and          | movie        | movi       |
| 3  | a      | a            | one          | one        |
| 4  | of     | of           | like         | like       |
| 5  | to     | to           | good         | time       |
| 6  | is     | is           | story        | good       |
| 7  | in     | in           | time         | stori      |
| 8  | that   | it           | great        | see        |
| 9  | I      | I            | well         | charct     |
| 10 | it     | that         | see          | make       |
| 11 | this   | s            | also         | well       |
| 12 | /><br  | this         | would        | get        |
| 13 | as     | as           | really       | great      |
| 14 | with   | with         | even         | watch      |
| 15 | was    | The          | much         | love       |
| 16 | for    | was          | first        | also       |
| 17 | but    | for          | people       | show       |
| 18 | his    | film         | get          | would      |
| 19 | the    | movie        | best         | realli     |
| 20 | on     | but          | love         | even       |

### Top 20 words for negative reviews

| No | Review | Clean Review | No Stopwords | Lemmatized |
|----|--------|--------------|--------------|------------|
| 1  | the    | the          | movie        | movi       |
| 2  | a      | a            | film         | film       |
| 3  | and    | and          | one          | one        |
| 4  | of     | to           | like         | like       |
| 5  | to     | of           | even         | make       |
| 6  | is     | is           | good         | bad        |
| 7  | in     | I            | bad          | even       |
| 8  | I      | in           | would        | get        |
| 9  | that   | it           | really       | time       |
| 10 | this   | that         | time         | good       |
| 11 | it     | this         | see          | charact    |
| 12 | /><br  | s            | story        | watch      |
| 13 | was    | was          | much         | would      |
| 14 | for    | movie        | get          | see        |
| 15 | with   | The          | people       | realli     |
| 16 | as     | for          | make         | look       |
| 17 | but    | with         | could        | stori      |
| 18 | movie  | t            | made         | scene      |
| 19 | The    | as           | plot         | act        |

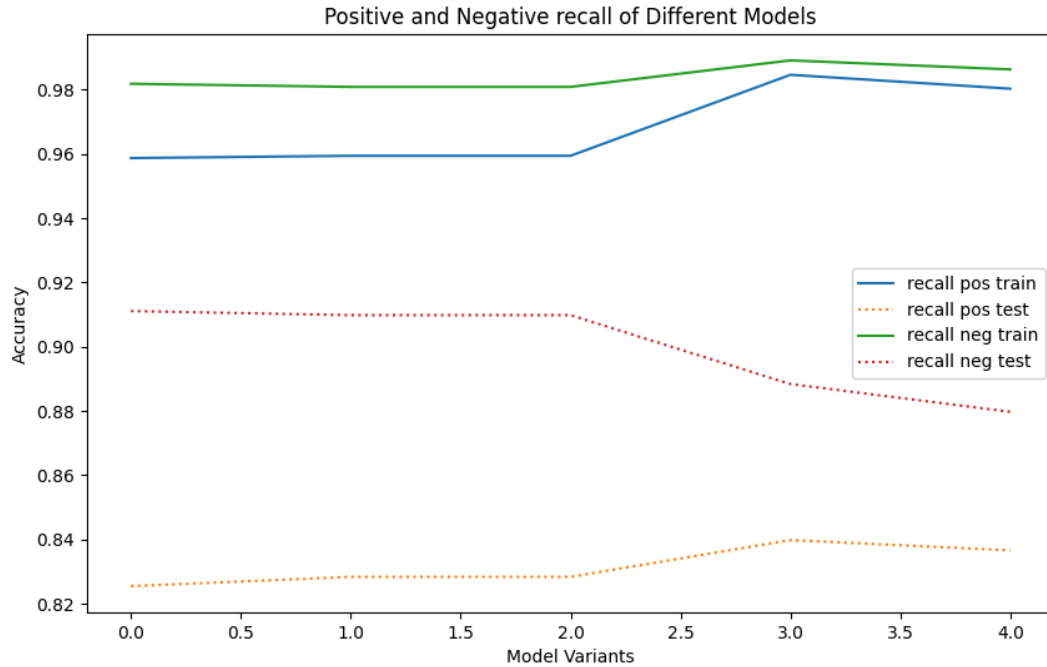
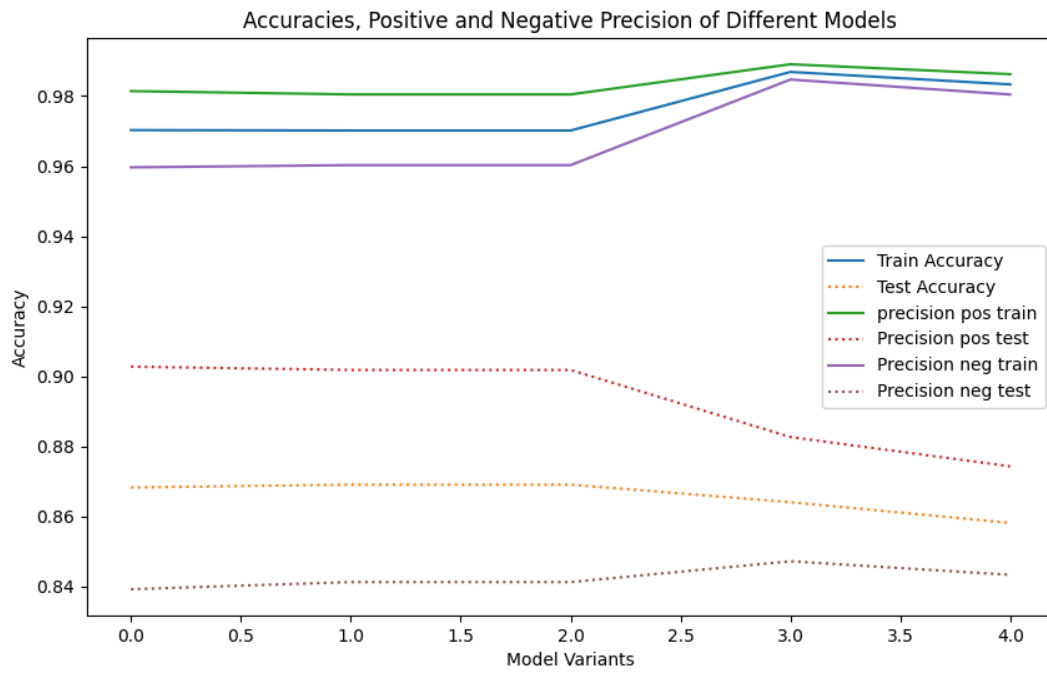
|    |    |      |      |      |
|----|----|------|------|------|
| 20 | on | film | well | much |
|----|----|------|------|------|

### **Step 7**

The lemmatize reduces words into their base forms, thus words with similar meanings would be taken as same by the model thus making model simpler and more performant

The lemmatization could have led to increase in word counts too for words like make which moved way up the list after lemmatization.

## **PART2**



## Analysis

Cleaning, removing stem words and lemmatization generally lead to improved accuracy, precisions and recall

However it can be noted that lemmatization leads to some huge drops for some metrics, especially recall for test data.

English when lemmatization and stemming are done can reduce the vocabulary complexity, making it easier for the model to understand and generate accurate predictions.

Using large datasets like IMDb can enhance model performance when lemmatization and stemming, as it provides a diverse range of language patterns and contexts for the model to learn from, resulting in more robust and accurate predictions.