Text Classification and Document Categorization

Bindeshwar Singh Kushwaha
PostNetwork Academy

Understanding Text with Machine Learning

- **Processing and understanding text allows extraction of meaningful information from raw data.**

Understanding Text with Machine Learning

- **Processing and understanding text allows extraction of meaningful information from raw data.**
- **Text data can be structured into features that machine learning algorithms can analyze.**

Understanding Text with Machine Learning

- **Processing and understanding text allows extraction of meaningful information from raw data.**
- **Text data can be structured into features that machine learning algorithms can analyze.**
- **Machine learning approaches include supervised, unsupervised, and deep learning techniques.**

Understanding Text with Machine Learning

- **Processing and understanding text allows extraction of meaningful information from raw data.**
- **Text data can be structured into features that machine learning algorithms can analyze.**
- **Machine learning approaches include supervised, unsupervised, and deep learning techniques.**
- **AI models combine data and algorithms to identify patterns and generate actionable insights.**

Text Classification and Categorization

- **Text classification organizes documents into predefined categories automatically.**

Text Classification and Categorization

- **Text classification organizes documents into predefined categories automatically.**
- **Each document is represented by features derived from its words or phrases.**

Text Classification and Categorization

- **Text classification organizes documents into predefined categories automatically.**
- **Each document is represented by features derived from its words or phrases.**
- **Applications include spam filtering, sentiment analysis, and news or topic categorization.**

Text Classification and Categorization

- **Text classification organizes documents into predefined categories automatically.**
- **Each document is represented by features derived from its words or phrases.**
- **Applications include spam filtering, sentiment analysis, and news or topic categorization.**
- **Effective classification relies on sufficient labeled data to train predictive models.**

Supervised vs. Unsupervised

- **Supervised: documents have labels (e.g., spam, non-spam).**

Supervised vs. Unsupervised

- **Supervised: documents have labels (e.g., spam, non-spam).**
- **Unsupervised: documents grouped by similarity without labels.**

Supervised vs. Unsupervised

- **Supervised: documents have labels (e.g., spam, non-spam).**
- **Unsupervised: documents grouped by similarity without labels.**
- **Document classification is a general problem applicable to many use cases.**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**
- **This involves:**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**
- **This involves:**
  - **Preprocessing text**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**
- **This involves:**
  - **Preprocessing text**
  - **Extracting features**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**
- **This involves:**
  - **Preprocessing text**
  - **Extracting features**
  - **Training a classification model**

Practical Example

- **Consider documents as sequences of sentences or paragraphs.**
- **Our task: classify each document into categories.**
- **This involves:**
  - **Preprocessing text**
  - **Extracting features**
  - **Training a classification model**
  - **Evaluating performance**

Concept of Text Classification

- **Represent documents as features (words, phrases, embeddings).**

Concept of Text Classification

- **Represent documents as features (words, phrases, embeddings).**
- **Use these features to assign documents to predefined categories.**

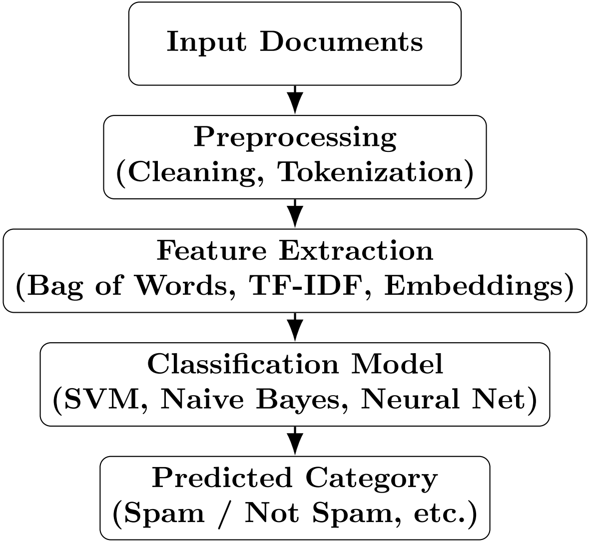
Concept of Text Classification

- **Represent documents as features (words, phrases, embeddings).**
- **Use these features to assign documents to predefined categories.**
- **Examples: spam detection, sentiment analysis, topic categorization.**

Concept of Text Classification

- **Represent documents as features (words, phrases, embeddings).**
- **Use these features to assign documents to predefined categories.**
- **Examples: spam detection, sentiment analysis, topic categorization.**
- **Process: Preprocessing $\rightarrow$ Feature Extraction $\rightarrow$ Classification.**

Simple Diagram of Text Classification



What are Features?

- **Features are unique, measurable attributes or properties of text data.**

What are Features?

- **Features are unique, measurable attributes or properties of text data.**
- **They can be absolute numeric values or categorical indicators.**

What are Features?

- **Features are unique, measurable attributes or properties of text data.**
- **They can be absolute numeric values or categorical indicators.**
- **Features represent text documents in a structured form for ML algorithms.**

What are Features?

- **Features are unique, measurable attributes or properties of text data.**
- **They can be absolute numeric values or categorical indicators.**
- **Features represent text documents in a structured form for ML algorithms.**
- **Example: words, phrases, or even TF-IDF scores.**

Feature Extraction

- **Process of converting raw text into structured features.**

Feature Extraction

- **Process of converting raw text into structured features.**
- **Usually involves transforming documents into numerical vectors.**

Feature Extraction

- **Process of converting raw text into structured features.**
- **Usually involves transforming documents into numerical vectors.**
- **Helps ML algorithms detect patterns and make predictions.**

Feature Extraction

- **Process of converting raw text into structured features.**
- **Usually involves transforming documents into numerical vectors.**
- **Helps ML algorithms detect patterns and make predictions.**
- **Challenge: finding the best way to represent unstructured text.**

Vector Space Model

- **A standard model to represent documents numerically.**

Vector Space Model

- **A standard model to represent documents numerically.**
- **Each document is represented as a vector of terms.**

Vector Space Model

- **A standard model to represent documents numerically.**
- **Each document is represented as a vector of terms.**
- **Terms can be words, tokens, or other features.**

Vector Space Model

- **A standard model to represent documents numerically.**
- **Each document is represented as a vector of terms.**
- **Terms can be words, tokens, or other features.**
- **Vector space enables algorithms to process and compare documents.**

Mathematical Formulation

- **Suppose there are n distinct words across all documents.**

Mathematical Formulation

- **Suppose there are n distinct words across all documents.**
- **Each document D can be represented as:**

$$D = (w_{1D}, w_{2D}, \ldots, w_{nD})$$

Mathematical Formulation

- Suppose there are n distinct words across all documents.
- Each document D can be represented as:

$$D = (w_{1D}, w_{2D}, \ldots, w_{nD})$$

- where w_{iD} = weight of word i in document D.

Mathematical Formulation

- Suppose there are n distinct words across all documents.
- Each document D can be represented as:

$$D = (w_{1D}, w_{2D}, \ldots, w_{nD})$$

- where w_{iD} = weight of word i in document D.
- Weight can be frequency, average occurrence, or TF-IDF.

Common Feature Extraction Models

- **Bag of Words (BoW)**

Common Feature Extraction Models

- **Bag of Words (BoW)**
- **TF-IDF (Term Frequency–Inverse Document Frequency)**

Common Feature Extraction Models

- **Bag of Words (BoW)**
- **TF-IDF (Term Frequency–Inverse Document Frequency)**
- **Advanced word vectorization models (Word2Vec, GloVe, BERT embeddings)**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

- **Negative Reviews:**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

- **Negative Reviews:**
- **I hated this film, it was boring and too long.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

- **Negative Reviews:**
- **I hated this film, it was boring and too long.**
- **The plot was weak and the acting was terrible.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

- **Negative Reviews:**
- **I hated this film, it was boring and too long.**
- **The plot was weak and the acting was terrible.**
- **Such a disappointing movie, I regret watching it.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- **I absolutely loved this movie. I loved the story and the characters.**
- **The acting was amazing and the story was touching.**
- **What a great experience. The visuals were great and memorable.**
- **This movie was thrilling and full of suspense.**
- **The cinematography and direction were excellent.**

- **Negative Reviews:**
- **I hated this film, it was boring and too long.**
- **The plot was weak and the acting was terrible.**
- **Such a disappointing movie, I regret watching it.**
- **This was the worst movie I have seen. Truly the worst experience.**

Movie Review Dataset (New 10 Sentences)

Positive Reviews:

- I absolutely loved this movie. I loved the story and the characters.
- The acting was amazing and the story was touching.
- What a great experience. The visuals were great and memorable.
- This movie was thrilling and full of suspense.
- The cinematography and direction were excellent.

- Negative Reviews:
- I hated this film, it was boring and too long.
- The plot was weak and the acting was terrible.
- Such a disappointing movie, I regret watching it.
- This was the worst movie I have seen. Truly the worst experience.
- Poor storyline and bad performances throughout.

BoW Vector Representation with Features and Class

Review	loved (x_1)	movie (x_2)	fantastic (x_3)	boring (x_4)	terrible (x_5)	great (x_6)	excellent (x_7)	worst (x_8)	acting (x_9)	story (x_{10})	Class
1	2	1	0	0	0	0	0	0	0	1	Positive
2	0	0	0	0	0	0	0	0	1	1	Positive
3	0	0	0	0	0	2	0	0	0	0	Positive
4	0	1	0	0	0	0	0	0	0	0	Positive
5	0	0	0	0	0	0	1	0	0	0	Positive
6	0	1	0	1	0	0	0	0	0	0	Negative
7	0	0	0	0	1	0	0	0	1	0	Negative
8	0	0	0	0	0	0	0	0	0	0	Negative
9	0	1	0	0	0	0	0	2	0	0	Negative
10	0	0	0	0	1	0	0	0	0	0	Negative

- **Each column x_i represents the count of the corresponding word in the review.**

BoW Vector Representation with Features and Class

Review	loved (x_1)	movie (x_2)	fantastic (x_3)	boring (x_4)	terrible (x_5)	great (x_6)	excellent (x_7)	worst (x_8)	acting (x_9)	story (x_{10})	Class
1	2	1	0	0	0	0	0	0	0	1	Positive
2	0	0	0	0	0	0	0	0	1	1	Positive
3	0	0	0	0	0	2	0	0	0	0	Positive
4	0	1	0	0	0	0	0	0	0	0	Positive
5	0	0	0	0	0	0	1	0	0	0	Positive
6	0	1	0	1	0	0	0	0	0	0	Negative
7	0	0	0	0	1	0	0	0	1	0	Negative
8	0	0	0	0	0	0	0	0	0	0	Negative
9	0	1	0	0	0	0	0	2	0	0	Negative
10	0	0	0	0	1	0	0	0	0	0	Negative

- **Each column x_i represents the count of the corresponding word in the review.**
- **Last column shows the class label (Positive/Negative).**

Use of Bag of Words in Classification

- **Positive words: {loved, fantastic, great, excellent}**

Use of Bag of Words in Classification

- **Positive words: {loved, fantastic, great, excellent}**
- **Negative words: {boring, terrible, worst, hated, poor}**

Use of Bag of Words in Classification

- **Positive words: {loved, fantastic, great, excellent}**
- **Negative words: {boring, terrible, worst, hated, poor}**
- **Words appearing multiple times indicate stronger sentiment (e.g., "loved" twice).**

Use of Bag of Words in Classification

- **Positive words: {loved, fantastic, great, excellent}**
- **Negative words: {boring, terrible, worst, hated, poor}**
- **Words appearing multiple times indicate stronger sentiment (e.g., "loved" twice).**
- **ML algorithms learn these patterns from training data.**

Use of Bag of Words in Classification

- **Positive words: {loved, fantastic, great, excellent}**
- **Negative words: {boring, terrible, worst, hated, poor}**
- **Words appearing multiple times indicate stronger sentiment (e.g., "loved" twice).**
- **ML algorithms learn these patterns from training data.**
- **New unseen reviews can be classified as Positive or Negative.**

Naive Bayes Classification

- **Naive Bayes is a probabilistic classifier based on Bayes' Theorem.**

Naive Bayes Classification

- **Naive Bayes is a probabilistic classifier based on Bayes' Theorem.**
- **Assumes conditional independence between features (words).**

Naive Bayes Classification

- **Naive Bayes is a probabilistic classifier based on Bayes' Theorem.**
- **Assumes conditional independence between features (words).**
- **Let $x = [x_1, x_2, \ldots, x_{10}]$ be the BoW vector of a review.**

Naive Bayes Classification

- **Naive Bayes is a probabilistic classifier based on Bayes' Theorem.**
- **Assumes conditional independence between features (words).**
- **Let** $x = [x_1, x_2, \ldots, x_{10}]$ **be the BoW vector of a review.**
- **Let** $y \in \{Positive, Negative\}$ **be the class label.**

Naive Bayes Classification

- **Naive Bayes is a probabilistic classifier based on Bayes' Theorem.**
- **Assumes conditional independence between features (words).**
- **Let $x = [x_1, x_2, \ldots, x_{10}]$ be the BoW vector of a review.**
- **Let $y \in \{Positive, Negative\}$ be the class label.**
- **Formula for predicting the class y:**

$$P(y \mid x) = \frac{P(y) \cdot P(x \mid y)}{P(x)}$$

Naive Bayes Classification

- Naive Bayes is a probabilistic classifier based on Bayes' Theorem.
- Assumes conditional independence between features (words).
- Let $x = [x_1, x_2, \ldots, x_{10}]$ be the BoW vector of a review.
- Let $y \in \{Positive, Negative\}$ be the class label.
- Formula for predicting the class y:

$$P(y \mid x) = \frac{P(y) \cdot P(x \mid y)}{P(x)}$$

- Since $P(x)$ is constant for all classes:

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{10} P(x_i \mid y)^{x_i}$$

Naive Bayes Classification

- Naive Bayes is a probabilistic classifier based on Bayes' Theorem.
- Assumes conditional independence between features (words).
- Let $x = [x_1, x_2, \ldots, x_{10}]$ be the BoW vector of a review.
- Let $y \in \{Positive, Negative\}$ be the class label.
- Formula for predicting the class y:

$$P(y \mid x) = \frac{P(y) \cdot P(x \mid y)}{P(x)}$$

- Since $P(x)$ is constant for all classes:

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{10} P(x_i \mid y)^{x_i}$$

- Here, $P(y)$ = prior probability of class y, $P(x_i \mid y)$ = likelihood of word x_i in class y.

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**
  - **"loved" appears 2 times**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**
  - **"loved" appears 2 times**
  - **"movie" appears 1 time**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**
  - **"loved" appears 2 times**
  - **"movie" appears 1 time**
  - **"story" appears 1 time**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**
  - **"loved" appears 2 times**
  - **"movie" appears 1 time**
  - **"story" appears 1 time**
  - **All other words in vocabulary = 0**

Naive Bayes Classification: Review 1 Word Counts

- **Review 1: "I absolutely loved this movie. I loved the story and the characters."**
- **Vocabulary: {loved (x_1), movie (x_2), fantastic (x_3), boring (x_4), terrible (x_5), great (x_6), excellent (x_7), worst (x_8), acting (x_9), story (x_{10})}**
- **Word counts in Review 1:**
  - **"loved" appears 2 times**
  - **"movie" appears 1 time**
  - **"story" appears 1 time**
  - **All other words in vocabulary = 0**
- **BoW vector representation: $x = [2, 1, 0, 0, 0, 0, 0, 0, 0, 1]$**

Naive Bayes: Compute Probability for Positive Class

- **Positive word counts (sum from all Positive reviews): loved=2, movie=3, story=2, fantastic=0, great=2, excellent=1, acting=1, boring=0, terrible=0, worst=0**

Naive Bayes: Compute Probability for Positive Class

- **Positive word counts (sum from all Positive reviews): loved=2, movie=3, story=2, fantastic=0, great=2, excellent=1, acting=1, boring=0, terrible=0, worst=0**
- **Total words in Positive reviews = 11**

Naive Bayes: Compute Probability for Positive Class

- **Positive word counts (sum from all Positive reviews): loved=2, movie=3, story=2, fantastic=0, great=2, excellent=1, acting=1, boring=0, terrible=0, worst=0**
- **Total words in Positive reviews = 11**
- **Probabilities for words in Review 1 (Positive):**

$$P(x_1 = 2 \mid Positive) = \frac{2}{11} \approx 0.182, \quad P(x_2 = 1 \mid Positive) = \frac{3}{11} \approx 0.273, \quad P(x_{10} = 1 \mid Positive) = \frac{2}{11} \approx 0.182$$

Naive Bayes: Compute Probability for Positive Class

- **Positive word counts (sum from all Positive reviews): loved=2, movie=3, story=2, fantastic=0, great=2, excellent=1, acting=1, boring=0, terrible=0, worst=0**
- **Total words in Positive reviews = 11**
- **Probabilities for words in Review 1 (Positive):**

$$P(x_1 = 2 \mid Positive) = \frac{2}{11} \approx 0.182, \quad P(x_2 = 1 \mid Positive) = \frac{3}{11} \approx 0.273, \quad P(x_{10} = 1 \mid Positive) = \frac{2}{11} \approx 0.182$$

- **Joint probability for Review 1 (Positive):**

$$P(x \mid Positive) \approx 0.182^2 \cdot 0.273 \cdot 0.182 \approx 0.00166$$

Naive Bayes: Compute Probability for Positive Class

- **Positive word counts (sum from all Positive reviews): loved=2, movie=3, story=2, fantastic=0, great=2, excellent=1, acting=1, boring=0, terrible=0, worst=0**
- **Total words in Positive reviews = 11**
- **Probabilities for words in Review 1 (Positive):**

$$P(x_1 = 2 \mid Positive) = \frac{2}{11} \approx 0.182, \quad P(x_2 = 1 \mid Positive) = \frac{3}{11} \approx 0.273, \quad P(x_{10} = 1 \mid Positive) = \frac{2}{11} \approx 0.182$$

- **Joint probability for Review 1 (Positive):**

$$P(x \mid Positive) \approx 0.182^2 \cdot 0.273 \cdot 0.182 \approx 0.00166$$

- **Multiply by prior $P(Positive) = 0.5$:**

$$P(Positive \mid x) \propto 0.5 \cdot 0.00166 \approx 0.00083$$

Naive Bayes: Compute Probability for Negative Class

- **Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2**

Naive Bayes: Compute Probability for Negative Class

- **Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2**
- **Total words in Negative reviews = 8**

Naive Bayes: Compute Probability for Negative Class

- **Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2**
- **Total words in Negative reviews = 8**
- **Probabilities for words in Review 1 (Negative):**

$$P(x_1 = 2 \mid Negative) = 0, \quad P(x_2 = 1 \mid Negative) = \frac{3}{8} = 0.375, \quad P(x_{10} = 1 \mid Negative) = 0$$

Naive Bayes: Compute Probability for Negative Class

- **Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2**
- **Total words in Negative reviews = 8**
- **Probabilities for words in Review 1 (Negative):**

$$P(x_1 = 2 \mid Negative) = 0, \quad P(x_2 = 1 \mid Negative) = \frac{3}{8} = 0.375, \quad P(x_{10} = 1 \mid Negative) = 0$$

- **Joint probability for Review 1 (Negative):**

$$P(x \mid Negative) = 0 \quad \text{(due to zero counts)}$$

Naive Bayes: Compute Probability for Negative Class

- **Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2**
- **Total words in Negative reviews = 8**
- **Probabilities for words in Review 1 (Negative):**

$$P(x_1 = 2 \mid Negative) = 0, \quad P(x_2 = 1 \mid Negative) = \frac{3}{8} = 0.375, \quad P(x_{10} = 1 \mid Negative) = 0$$

- **Joint probability for Review 1 (Negative):**

$$P(x \mid Negative) = 0 \quad \text{(due to zero counts)}$$

- **Multiply by prior $P(Negative) = 0.5$:**

$$P(Negative \mid x) = 0$$

Naive Bayes: Compute Probability for Negative Class

- Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2
- Total words in Negative reviews = 8
- Probabilities for words in Review 1 (Negative):

$$P(x_1 = 2 \mid Negative) = 0, \quad P(x_2 = 1 \mid Negative) = \frac{3}{8} = 0.375, \quad P(x_{10} = 1 \mid Negative) = 0$$

- Joint probability for Review 1 (Negative):

$$P(x \mid Negative) = 0 \quad \text{(due to zero counts)}$$

- Multiply by prior $P(Negative) = 0.5$:

$$P(Negative \mid x) = 0$$

- Compare: $P(Positive \mid x) > P(Negative \mid x)$

Naive Bayes: Compute Probability for Negative Class

- Negative word counts (sum from all Negative reviews): loved=0, movie=3, story=0, fantastic=0, great=0, excellent=0, acting=1, boring=1, terrible=1, worst=2
- Total words in Negative reviews = 8
- Probabilities for words in Review 1 (Negative):

$$P(x_1 = 2 \mid Negative) = 0, \quad P(x_2 = 1 \mid Negative) = \frac{3}{8} = 0.375, \quad P(x_{10} = 1 \mid Negative) = 0$$

- Joint probability for Review 1 (Negative):

$$P(x \mid Negative) = 0 \quad \text{(due to zero counts)}$$

- Multiply by prior $P(Negative) = 0.5$:

$$P(Negative \mid x) = 0$$

- Compare: $P(Positive \mid x) > P(Negative \mid x)$
- Conclusion: Review 1 is classified as Positive.

Reach PostNetwork Academy

Website

www.postnetwork.co

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy

Reach PostNetwork Academy

Website

www.postnetwork.co

YouTube Channel

www.youtube.com/@postnetworkacademy

Facebook Page

www.facebook.com/postnetworkacademy

LinkedIn Page

www.linkedin.com/company/postnetworkacademy

GitHub Repositories

www.github.com/postnetworkacademy

Thank You!