

Machine Learning in Text Classification

Syed Muhammad Faraz Ali (19K-0861)
Muhammad Faraz Ashraf (19K-0957)

National University of Computer and Emerging Sciences

22th April 2020

Table of contents

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

1 Background

- What is Machine Learning
- Requirements to Implement Machine Learning
- Machine Learning in NLP

2 Implementation

- Feature Extraction
- Feature Usage in Model
- Machine Learning Models
- Model Evaluation Approach and Metrics

Table of contents

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

1 Background

- What is Machine Learning
- Requirements to Implement Machine Learning
- Machine Learning in NLP

2 Implementation

What is Machine Learning

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning

Requirements

NLP

Implementation

Feature Extraction

Feature Usage

ML Models

Evaluation

Machine learning is the field of computer science which give the machine an ability to learn from data without being **explicitly** programmed.

What is Machine Learning

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

Machine learning is the field of computer science which give the machine an ability to learn from data without being **explicitly** programmed.

- Supervised Learning

- 1 Classification (Categorical or Fixed-Discrete Target)
- 2 Regression (Continuous Target)

What is Machine Learning

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

Machine learning is the field of computer science which give the machine an ability to learn from data without being **explicitly** programmed.

- Supervised Learning

- 1 Classification (Categorical or Fixed-Discrete Target)
- 2 Regression (Continuous Target)

- Unsupervised Learning

- 1 Clustering
- 2 Dimensionality Reduction

What is Machine Learning

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

Machine learning is the field of computer science which give the machine an ability to learn from data without being **explicitly** programmed.

- Supervised Learning

- 1 Classification (Categorical or Fixed-Discrete Target)
- 2 Regression (Continuous Target)

- Unsupervised Learning

- 1 Clustering
- 2 Dimensionality Reduction

- Reinforcement Learning

NOTE: Rule-based engine and rote-learner are not machine learning

Requirements to Implement Machine Learning

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning

Requirements

NLP

Implementation

Feature Extraction

Feature Usage

ML Models

Evaluation

- 1 Question that needs to be answered
- 2 Availability of Data
- 3 Selection of Features
- 4 Selection of ML Models
- 5 Performance Metric

Machine Learning in NLP

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning

Requirements

NLP

Implementation

Feature Extraction

Feature Usage

ML Models

Evaluation

- Supervised Learning
 - 1 Document Classification such as spam filter
 - 2 Annotation such as Named Entity Recognition
 - 3 Sentiment Analysis
- Unsupervised Learning
 - 1 Topic Modeling
 - 2 Document Clustering
 - 3 Reducing Dimension of Data

Table of contents

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

1 Background

2 Implementation

- Feature Extraction
- Feature Usage in Model
- Machine Learning Models
- Model Evaluation Approach and Metrics

Feature Extraction

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

Usually NLP data is in document form and it is unstructured. Therefore, relevant features need to be extracted from them.

- Count Vectorizer
- TF-IDF Vectorizer
- Binarized Bag-of-words
- N-grams of words

Feature Usage in Model

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

The features extracted can be used in two ways in a machine learning model.

- as-is
- Dimensionality Reduction
 - 1 Principal Component Analysis (PCA)
 - 2 Singular Value Decomposition (SVD)
 - 3 Random Projection

Usually the matrix returned by feature extraction is a sparse-matrix. Performing PCA can be computationally expensive on a sparse-matrix. SVD and Random Projection are preferable.

Machine Learning Models

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning

Requirements

NLP

Implementation

Feature Extraction

Feature Usage

ML Models

Evaluation

■ Supervised Learning

- 1 Naive Bayes
- 2 Logistic Regression
- 3 Random Forest
- 4 Support Vector Machine (SVM)
- 5 Neural Networks/ Deep Learning

■ Unsupervised Learning

- 1 Latent Dirichlet Allocation (LDA)
- 2 Latent Semantic Indexing (LSI)
- 3 Matrix Factorization

Model Evaluation Approach and Metrics

Natural
Language
Processing

Faraz Ali
Faraz Ashraf

Background

Machine Learning
Requirements
NLP

Implementation

Feature Extraction
Feature Usage
ML Models
Evaluation

■ Evaluation Approach

- 1 Hold-out
- 2 Cross Validation
- 3 Leave one out Cross Validation

■ Evaluation Metrics (Classification)

- 1 Accuracy
- 2 Precision
- 3 Recall
- 4 F1-score

THANK YOU