LDA_Classifiy Class Documentation

1. Load_models (lda_model_path, dictionary_model_path)
- Receives LDA saved model path.
- Receives generated Dictionary save path.
- Loads generated dictionary at training.
- Passes addresses to load_LDA + lod_dict methods respectively.

2. Tokenize (document)

-preforms tokenization on each individual document.

- splits document into sentences
- sentences into words.
- remove punctuation.
- Words that have fewer than 3 characters are removed.
- All stop words are removed.

- passes each tokenized word(token) to lemmatize method.

- passes each lemmatized word(token) to stemmed method.

- appends processed words in a new document.

- yields a document of processed (tokenized + stemmed + lemmatized) words.

3. Lemmatize (word)

-preforms lemmatization on each passed token(word).

- words in third person are changed to first person.
- verbs in past and future tenses are changed into present sentences into words.

- yields lemmatized token.

4. Stemm (word)

-preforms stemming on each passed token(word).

- words are reduced to their root form

- yields stemmed token.

5. Get_topics ( )
- Loads generated topics from trained LDA model.
- Each topic is composed of a distribution of specific words.

6. Load_LDA (save_path)
- Loads pre-trained (saved) LDA model.

7. Load_dict (save_path)

- Loads pre-calculated (saved) dictionary.

8. classify (documents)

- Preforms preprocessing (tokenization, etc..) on new unseen documents.
- Generates bag of words for unseen documents using pre-generated dictionary.
- Passes unseen document to pre-trained LDA model that Generates topics with corresponding weights.