✓ **Congratulations! You passed!**

Grade received 100%   To pass 75% or higher

[**Go to next item**]

Feature extraction from text and images

Total points 4

1. TF-IDF is applied to a matrix where each column represents a word, each row represents a document, and each value shows the number of times a particular word occurred in a particular document. Choose the correct statements.

   1 / 1 point

   ☑ IDF scales features inversely proportionally to a number of word occurrences over documents

   ✓ **Correct**
   Correct! Purpose of IDF is to decrease importance of most frequent words.

   ☐ IDF scales features proportional to the frequency of word's occurrences

   ☐ TF normalizes sum of the column values to 1

   ☑ TF normalizes sum of the row values to 1

   ✓ **Correct**
   Correct!

2. What of these methods can be used to preprocess texts?

   1 / 1 point

   ☑ Stemming

   ✓ **Correct**
   Correct! Stemming is used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

   ☐ Plumping

   ☑ Lowercase transformation

   ✓ **Correct**
   Correct! Lowercase is used to unite words written in UPPERCASE and lowercase.

   ☑ Lemmatization

   ✓ **Correct**
   Correct! Lemmatization is used to reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

   ☐ Plumbing

   ☐ Levenshteining

   ☑ Stopwords removal

   ✓ **Correct**
   Correct! After feature generation stopwords result in useless features that interfere with useful ones.

3. What is the main purpose of Lemmatization and Stemming?

   1 / 1 point

   ◉ To reduce inflectional forms and sometimes derivationally related forms of a word to a common base form.

   ○ To reduce significance of common words.

   ○ To remove words which are not useful.

   ○ To induce common word amplification standards to the most useful for machine learning algorithms form.

   ✓ **Correct**
   You are right, this is the exact purpose of Lemmatization and Stemming!

4. To learn Word2vec embeddings we need ...

   1 / 1 point

Text corpora

☐ Labels for the documents in the corpora

☐ Labels for each word in the documents in the corpora

☐ GloVe embeddings