**AYDIN ADNAN MENDERES UNIVERSITY**

**ENGINEERING FACULTY**

**COMPUTER SCIENCE ENGINEERING DEPARTMENT**



**Finding Mean Values from 6 Word2Vec Models**

**CSE431 – Natural Language Processing with Machine Learning 2023/2024**

**Burak TÜZEL**

**Talha Alper ASAV**

**Lecturer:**

**Asst. Prof. Dr. Fatih SOYGAZİ**

**Finding Mean Values from 6 Word2Vec Models**

**Installing necessary environment and importing the libraries:**

1. A screenshot of a computer program

   Description automatically generatedInstall Jupyter Notebook

**Downloading the Latest Wiki Dump and Training Our Model:**

Here is the wiki dump link to download latest dump: <https://dumps.wikimedia.org/trwiki/>



At the end we saved our model for later. So we can use it later on without training again.

**Here Is Some Operations We Did to Test Our Model:**

A group of colorful text

Description automatically generatedIn here we tried different ways to test our model. It can find for example (king+women)-man = queen

**Loading Other 5 Models We Are Going to Use:**

Here we loaded the “2018 trmodel” and other 4 that comes with gensim library.

A screenshot of a computer program

Description automatically generated

**Finding The Similar Words with The Given (‘Spor’,’Siyaset’,’Magazin’,’Ekonomi’) Words:**



**An Example of The Found Similar Words:**

These words were created from the model we trained with the Wiki Dump.

A screenshot of a computer code

Description automatically generated

**Appending the Found Similar Words:**

Here we appended the words we found from each model. So we can use them while calculating the mean values with lexicons.

A close-up of a computer code

Description automatically generated

**Loading The Lexicons We Are Going to Use:**

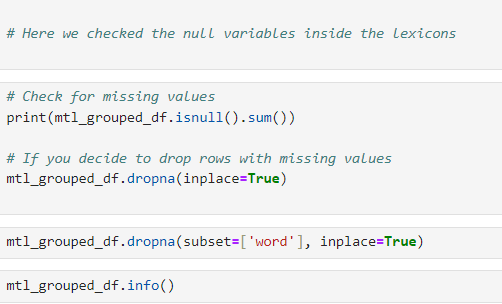
These are the lexicons. NRC-VAD only has “Valence” “Arousal” and “Dominance” values while the

MTL-Grouped has additional emotion values such as “Joy” “Anger” “Sadness” etc.



**Removing NaN and Checking Datatypes of Lexicons:**

Here we removed the nan inside the MTL-Grouped lexicon.



And here we checked the datatypes of lexicons.

**Here We Calculated the Mean Values of Similar Words and Created Tables:**

Here we found mean values for Turkish Models.

A screenshot of a computer program

Description automatically generated

A screenshot of a computer screen

Description automatically generatedAnd here is the rest of the models that are in English.

A screenshot of a computer

Description automatically generated**Here Are Some Examples of the Results of Each Model With Given Test Words And Their Similar Words:**

A screenshot of a computer

Description automatically generatedHere is another examples that comes from Glove Vectors Twitter 200.

**REFERENCES**

1. <https://github.com/akoksal/Turkish-Word2Vec/wiki/>
2. <https://dumps.wikimedia.org/trwiki/>
3. <https://drive.google.com/drive/folders/1IBMTAGtZ4DakSCyAoA4j7Ch0Ft1aFoww>
4. <https://saifmohammad.com/WebDocs/Lexicons/NRC-VAD-Lexicon.zip>
5. <https://zenodo.org/record/3756607/files/MTL_grouped.zip?download=1>

ssss