ÇANKAYA ÜNİVERSİTESİ
BİLGİSAYAR MÜHENDİSLİĞİ

MULTI LABEL CLASSIFICATION OF NEWS TEXT

Team Members
Can Koral ADALI 201411002
Bihter ÖZUÇAK 201411048
Miray PADIR 201411050
Kardelen YILDIRIM 201311060

Advisor
Prof. Dr. Erdoğan DOĞDU

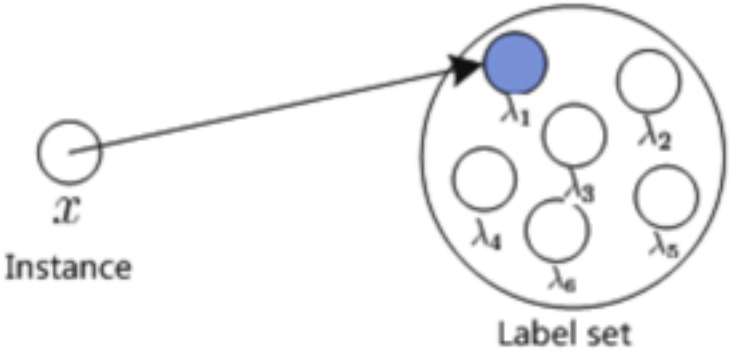
Co-Advisor
Asst. Prof. Dr. Roya CHOUPANI

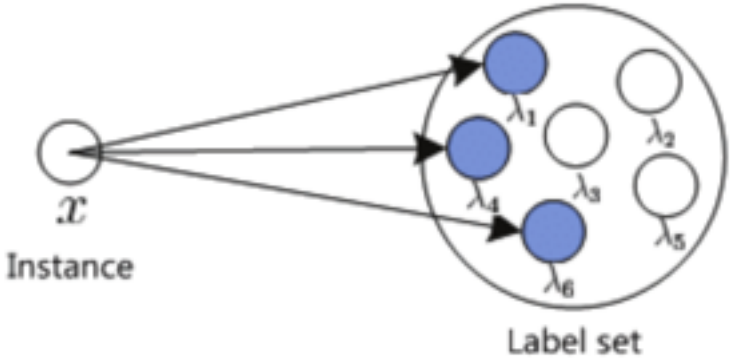
PROJECT DEFINITION

A website that can multi-label the given Turkish text documents by analyzing them with deep learning methods.

- Multi-labeling the text document by given model.
- Compatible with every web browser that can run JavaScript.

Multi-class and Multi-label





Maın features

- Multi-label the given text document(s) in a short time.
- Accuracy on given labels regarding success on models.
- It can analyze one or many given text documents depending on upload.

PROBLEM

- Many digital documents everywhere over the internet.
- **Manual labeling** with subjects (labels) is difficult, even impossible.
- An intelligent tool is needed for **automatic labeling**.

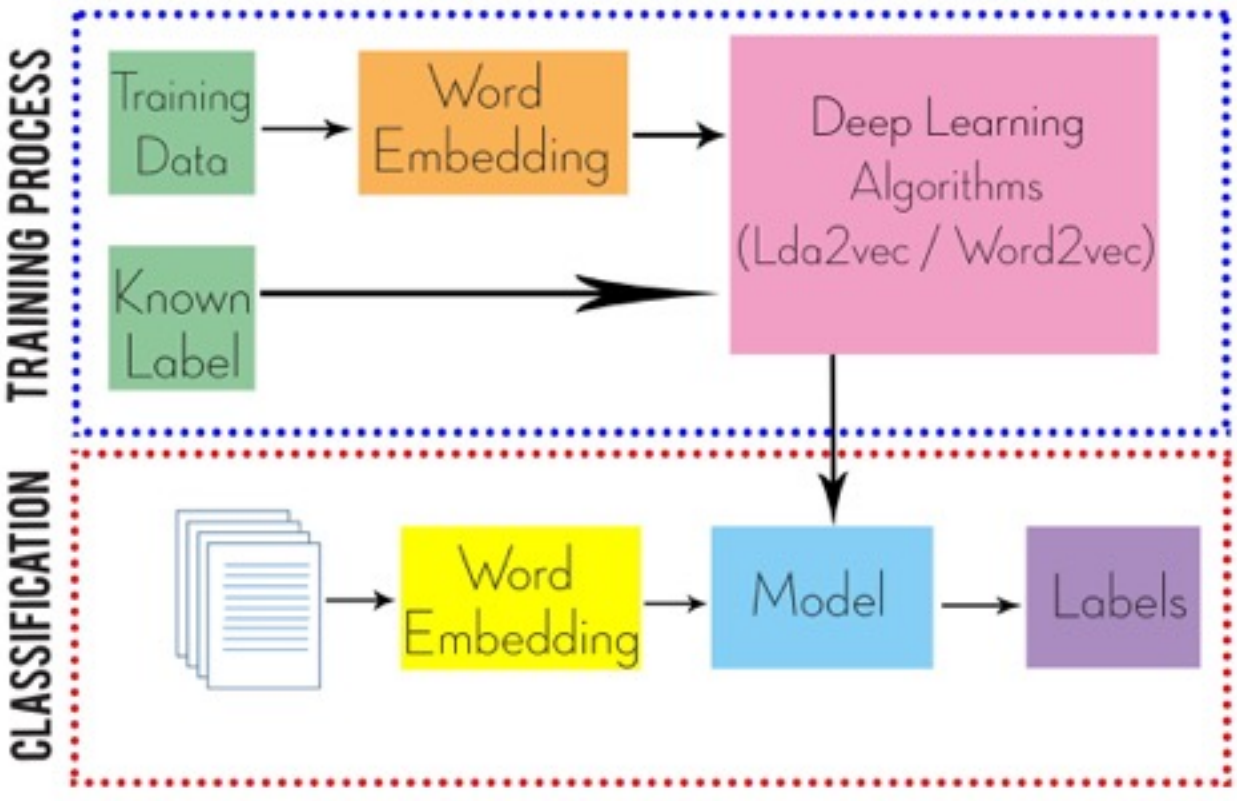
METHODS & TOOLS

- We applied deep learning methods. Such as **Word2vec** with using **Python**.

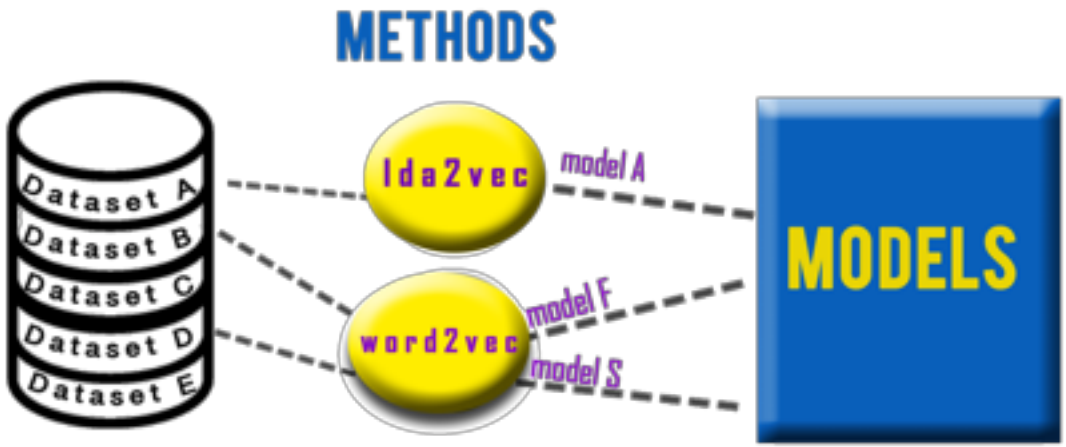
- System used given news datasets for deep learning methods. So that, system can use it and create models.

- System tried and found the best model for the given news text document(s) so that it can give better accuracy for labels to classified document(s).

What was our proposal?




Dataset


METHODS
Dataset A
Dataset B
Dataset C
Dataset D
Dataset E
lda2vec
model A
word2vec
model F
model S
MODELS

OUR SOLUTION

- Our solution based on two processes. One of them is train model based on given datasets. The other one is using this trained models on given news text documents.
- Training is performed by using Magpie library which is using Google's, Tensorflow and Keras modelling algorithm.

REFERENCES

1. Yaghoobzadeh, Y., Kann, K., & Schütze, H. (2018). Evaluating Word Embeddings in Multi-label Classification Using Fine-grained Name Typing. arXiv preprint arXiv:1807.07186
2. A. Clare and R. D. King, “Knowledge Discovery in Multi-Label Phenotype Data,” in: Lecture Notes in Computer Science. Springer,2001.
3. Auria, L., & Moro, R. A. (2008). Support vector machines (SVM) as a technique for solvency analysis.
4. Mohammad, A. H., Alwada’n, T., & Al-Momani, O. (2018). Arabic text categorization using support vector machine, Naïve Bayes and neural network. GSTF Journal on Computing (JoC), 5(1).
5. van Meeuwen, F. W. (2013). Multi-label text classi? cation of news articles for ASDMedia (Master's thesis).

ACKNOWLEDGEMENT

- Prof. Dr. Erdoğan DOĞDU

- Asst. Prof. Dr. Roya CHOUPANI

THANK YOU FOR LISTENING