# SENTIMENT ANALYSIS OF TURKISH POLITICAL COLUMNS WITH TRANSFER LEARNING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MESUT KAYA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
COMPUTER ENGINEERING

AUGUST 2013

Approval of the thesis:

**SENTIMENT ANALYSIS OF TURKISH POLITICAL COLUMNS WITH TRANSFER LEARNING**

submitted by **MESUT KAYA** in partial fulfillment of the requirements for the degree of **Master of Science in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Adnan Yazıcı
Head of Department, **Computer Engineering** _____

Prof. Dr. Ismail H. Toroslu
Supervisor, **Computer Engineering Department, METU** _____

Dr. Guven Fidan
Co-supervisor, **Information Systems Department, METU** _____

**Examining Committee Members:**

Assoc. Prof. Dr. Pınar Karagöz
Computer Engineering Department, METU _____

Prof. Dr. Ismail H. Toroslu
Computer Engineering Department, METU _____

Dr. Guven Fidan
Information Systems Department, METU _____

Assoc. Prof. Dr. Tolga Can
Computer Engineering Department, METU _____

Assoc. Prof. Dr. Osman Abul
Computer Engineering Department, TOBB ETÜ _____

**Date:** _____

**I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.**

Name, Last Name: MESUT KAYA

Signature :

# ABSTRACT

SENTIMENT ANALYSIS OF TURKISH POLITICAL COLUMNS WITH TRANSFER LEARNING

KAYA, Mesut

M.S., Department of Computer Engineering

Supervisor : Prof. Dr. Ismail H. Toroslu

Co-Supervisor : Dr. Guven Fidan

August 2013, 49 pages

Sentiment Analysis aims to determine the attitude (sense, emotion, opinion etc.) of a speaker or a writer with respect to a specified topic by automatically classifying a textual data.

With the recent explosive growth of the social media content on the Web, people post reviews of products on merchant sites and express their views about almost anything in their personal blogs, pages at social network sites like Facebook, Twitter, and Blogger. Therefore, sentiment analysis has become a major area of interest in the field of NLP.

Up to date, most of the research carried on sentiment analysis was focused on highly subjective English short texts, such as product or movie reviews. In this thesis, sentiment classification is applied on Turkish political columns. Both sentiment analysis on news domain and Turkish language received less attention. Besides, in order to reduce the expense of collection and annotation of the data, Transfer Learning, which is recently adopted to sentiment classification tasks, mechanisms are applied to sentiment analysis of Turkish political columns.

Keywords: Sentiment Analysis, Machine Learning, News Domain, NLP, Transfer Learning, Turkish

# ÖZ

TÜRKÇE POLİTİK KÖŞE YAZILARININ SENTİMENT ANALİZİ

KAYA, Mesut

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. Ismail H. Toroslu

Ortak Tez Yöneticisi : Dr. Guven Fidan

Ağustos 2013 , 49 sayfa

Düşünce Analizi ya da Düşünce Çözümleme bir konuşmacının ya da bir yazarın belirli bir konu hakkında görüşünün (algı, duygu, düşünce vs.) otomatik olarak sınıflandırılarak belirlenmesidir.

İnternetteki sosyal medya içeriğinin son zamanlarda aşırı büyümesiyle birlikte, internet kullanıcıları ticari web sitelerindeki ürünler hakkında yorumlar yapıp hemen hemen her şey hakkında düşüncelerini kişisel bloglarında, Facebook, Twitter ve Blogger gibi sosyal ağ sitelerinde belirtmektedirler. Böylece, düşünce analizi Doğal Dil İşleme alanında önemli bir araştırma alanı oluşturmuştur.

Günümüze kadar yapılmış çoğu düşünce analizi araştırmaları İngilizce subjektif kısa ürün ya da film yorumları verisine odaklanmıştır. Bu tez kapsamında, düşünce analizi Türkçe politik köşe yazılarına uygulanmıştır. Düşünce analizi çalışmalarında hem Türkçe metinler hem de politik köşe yazıları daha az çalışma konusu olmuştur. Düşünce analizi çalışmalarında önemli bir çalışma gerektiren veri toplama ve etiketleme işlemlerine gereksinimi azaltmak için Transfer Learning, düşünce analizi alanında yeni yeni kullanılmakta olan bir araştırma alanı, mekanizmaları düşünce analizi metodlarına entegre edilmiştir.

Anahtar Kelimeler: Düşünce Analizi, Makinalı Öğrenme, Haber Alanı, Doğal Dil İşleme, Transfer learning, Türkçe

*To my family...*

# ACKNOWLEDGMENTS

I would like to express my sincere appreciations to my advisor, Prof. Dr. Ismail Hakkı Toroslu and my co-advisor Dr. Güven FİDAN for their guidance, support, encouragement and positive attitudes throughout my study. I am gladful for having chance of working with them.

I want to thank my friends who supported me morally and helped me a lot.

I would like to thank my company AGMLab for supporting my thesis, both for research part and data collection, annotation parts.

I am very grateful that my family encouraged me throughout the study.

Although I do not know many of them, my Twitter account followers were always with me throughout my study. I want to express a special thanking to them.

# TABLE OF CONTENTS

# LIST OF TABLES

TABLES

# LIST OF FIGURES

FIGURES

# LIST OF ALGORITHMS

ALGORITHMS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AI | Artificial Intelligence |
| HTML | Hyper Text Markup Language |
| KNN | K-nearest Neighbor |
| ME | Maximum Entropy |
| ML | Machine Learning |
| NB | Naive Bayes |
| NER | Named Entity Recognition |
| NLP | Natural Language Processing |
| POS | Part opf Speech |
| SCL | Structural Correspondence Learning |
| SVM | Support Vector Machines |
| TF-IDF | Term Frequemcy - Inverse document Frequency. |

# CHAPTER 1

# INTRODUCTION

## 1.1 Motivation

The Internet has become a global forum where people express their opinions. With the recent explosive growth of the social media content on the Web, people post reviews of products on merchant sites and express their views about almost anything in their personal blogs, pages at social network sites like Facebook, Twitter and Blogger. According to a recent study[1]there are 1 billion monthly active users on Facebook, 200 million monthly active users on Twitter and throughout 2012 every day 175 million tweets are sent in average every day. Looking only the statistics throughout 2012 for Facebook and Twitter, the amount of data is enormous and the amount of subjective data available on the web is very large, too. People are also eager to know what individual journalists and columnists are thinking or feeling about subjects such as politicians, political parties and social issues. With the rapid growth of Twitter among people almost all of the columnists and journalists have verified Twitter accounts. Therefore, it is possible to analyze columnists' opinions and feelings, too.

Using the enormous data available on the Web it is possible to analyze them automatically with natural language processing (NLP) techniques, information retrieval (IR), data mining methods etc. One of the approaches to analyze textual data is sentiment analysis, aiming to extract feelings from a given text. Sentiment analysis become an important research field and is applied to different domains by using different approaches.

Figure 1.1 shows Worl Wide Web Search Interest of "Sentiment Analysis" from 2004 to present by using Google Trends[2]. It can be observed that with the growth of the available data on the internet, the need for sentiment analysis also increases. Various statistical and linguistic methods have been developed for the sentiment analysis of English texts for different domains. In the case of Turkish and other morphologically rich languages, however, sentiment analysis is a new field of research and not much work has been published in the field. Therefore, in this thesis the effectiveness of using machine learning techniques on Turkish sentiment analysis of news data is examined (analyzing columns of Turkish columnists to understand whether they support or criticize a political party, politician, or social issue). Since there were no tagged news data available, collected political columns are tagged.

One of the difficulties of the sentiment classification of news data is the lack of the tagged data and since the columns are not short texts, the annotation job is difficult and expensive.

[1]http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/

[2]http://www.google.com/trends/explore#q=sentiment%20analysis&cmpt=q

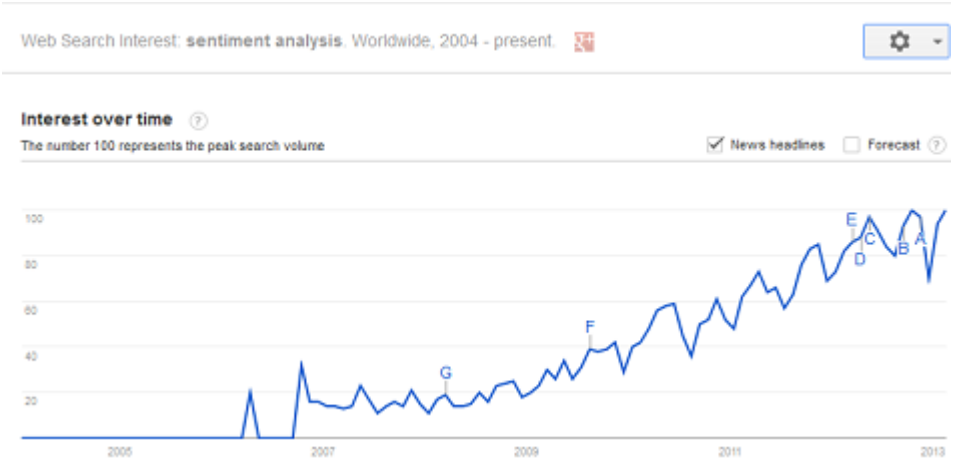


Figure 1.1: Web Search İnterest "Sentiment Analysis" WorldWide from 2004 - present by Google Trends

In order to overcome the tagged data problem, “Transfer Learning”, aiming to extract the knowledge from source tasks to be applied to a target task, is adapted [29]. In the scope of the thesis, a novel approach to transfer learning is covered.

The aim of this thesis is applying sentiment classification methods to Turkish political news data, which is an initial study, and to improve the performance of classification methods covering Transfer Learning methods, trying a new method.

## 1.2 Sentiment Analysis

Term "Sentiment" is related with feelings, attitudes, emotions and opinions which are not facts but subjective impressions.

Sentiment Analysis, sometimes referred to as Opinion Mining also aims to identify, extract or characterize the sentiment of a given text by using NLP, statistics or machine learning (ML) methods.

The sentiment, semantic orientation or polarity of a given text is generally stated as a binary opposition like:

- positive/negative
- good/bad
- for/against
- like/dislike

Pang and Lee (2008) [7] states that there are different applications that use sentiment analysis in different domains like:

- Applications to review-related websites.
- Applications as a sub-component technology like recommendation systems, flame detection, question answering systems, summarization, citation analysis.
- Applications in business and government intelligence.
- Applications across different domains like politics, sociology etc.

Having such a wide application area and domains, our motivation is to use sentiment analysis in politics, by using Turkish political news domain.

## 1.3 News Domain

For our research we choose to work in the news domain, in the sentiment classification of political columns. Our motivation for applying sentiment analysis to news data is that there is not much previous work in the area. Most of the studies use reviews (movie reviews, product reviews etc.), since the writers often summarize their sentiments in their reviews. Besides, it becomes easy to annotate training data as positive or negative since most of the web sites accepting user reviews also provide a rating system, i.e. a 0-5 scale star system. Moreover, reviews are short and relatively easy to analyze.

Balahur and Steinberger (2009) [2] states that for news articles, although support or criticism is sometimes expressed, journalists or columnists express their opinions indirectly. In most news data, journalists use a clear language to give an impression of objectivity, and for the same reason they state their opinions indirectly by embedding statements in a more complex discourse or argument structure [3]. The main reason why we choose columnists' work is that, unlike most of the news data, in their columns they use subjective language to express their opinions. This results in either support or criticism about *a politician, a political party, a social issue* and so on. However, there are still some difficulties and problems encountered in the task of the sentiment classification of political columns.

Difficulty with annotating political columns as positive or negative is a challenging job since in the political news domain a sentence can be positive for one annotator but negative for another. For example, consider the following Turkish sentence:

*"Oysa 'A partisi' boyun eğmeye hiç niyetli görünmüyor."* (But, 'Party A' does not seem to give up.).

This may be annotated as positive by an annotator supporting the political views of 'Party A', and negative by another annotator not supporting the party's views.

Another difficulty with using political columns is that columnists most of the time put positive criticism and negative criticism together. For example, a columnist supporting the political views of 'Party A' rather than 'Party B', writes positive criticism about 'Party A' and negative criticism about 'Party B' in the same column. Therefore, while annotating the columns we look at the overall sentiment of the articles.

Concerning the difficulties explained, we tried to collect columns demonstrating positive (support) or negative criticism about a person, political party, popular topic etc.

## 1.4 Outline of the Thesis

In Chapter 2, literature survey of related work on Sentiment Analysis and Transfer Learning is given. Chapter 3 covers detailed background information about different methods and techniques used in this thesis. In Chapter 4, data sets used in the experiments are briefly covered and the details of the experimental setup are given. Experimental results and the discussion of the results are given in Chapter 5. In Chapter 6, the thesis is summarized and conclusions are given. In the "Future Work" section what possible further improvements can be studied in the future are discussed. Finally, in the Appendix part, sample Turkish and English political columns are given.

# CHAPTER 2

# LITERATURE SURVEY

Early research on sentiment classification of documents is based on cognitive linguistics models [18], [34], or on the construction of discriminant-word lexicons [20], [13], [39].

Some research on sentiment classification focuses on classifying the semantic orientation of words or phrases [17], [41].

Research that aims at distinguishing the author's polarity on certain topics from document level [40], [30], [14] to sentence level [19], [24] has been performed.

In the recent studies, subjectivity extraction has become the area of interest. Sentence-level subjectivity classifiers are trained and it is shown by using selected subjective sentences that only sentiment analysis gets better quality results [43].

## 2.1 Sentiment Analysis With Machine Learning

Mostly "supervised learning" is applicable to sentiment classification as a machine learning approach. Supervised learning includes text classification tasks. A number of ML techniques are used in sentiment classification of texts like Naive Bayes (NB), maximum entropy (ME), support vector machines (SVM) N-Gram language model and K-nearest neighbor (KNN).

Pang et al. (2002) [30] have one of the major research in sentiment classification of textual data with ML techniques. They have used movie reviews and compared the performance of NB, ME and SVM by using different features like unigrams, bigrams, combination of both together with part of speech (POS) information. Their experiments show that SVM performs better than the other two methods.

Tan and Zhang (2008) [35], to find out the best ML technique on sentiment classification of Chinese documents, compared performance of SVM, NB and KNN. They have shown that SVM outperforms other techniques.

There are also studies that use the combination of rule-based classification, supervised learning and machine learning. For example, Prabowo and Thellwall (2009) [32] carried out a hybrid classification, namely if one classifier fails to classify the document the next classifier tries to classify it. This approach is known as a hybrid classification for sentiment analysis.

Another combination method is to combine the outputs of some base classifiers to have a last

integrated output. Xia et al. (2011) [33] used NB, ME and SVM as base classifiers to form an integrated output by using different feature sets.

Most of the researches available in the literature show that SVM outperforms the other techniques in sentiment analysis. However, there are some cases that surprisingly NB outperforms SVM. For instance, Zhang et al. (2011) [45] focused on a written variety of Chinese known as Cantonese and show that NB performs better than SVM.

Although most of the researches on sentiment analysis that use ML use techniques like NB, ME, SVM etc., there are some recent studies that use neural networks in sentiment classification [44], [26].

## 2.2 Sentiment Analysis in News Domain

Most previous works focus on the sentiment analysis of highly subjective texts, such as product or movie reviews. In this work we try to focus on the sentiment classification of columnists' work which is in news domain and has received less attention, although some initial work on sentiment analysis in the news domain has been conducted recently [16], [4]. Mihelcea and Strapparava [37] tried to classify newspaper titles according to their emotion, and Godbole et al. [28] presented a system to assign scores indicating positive or negative opinion to each distinct entity in a text corpus collected from blogs and news.

One of the most major pieces of research on sentiment classification of news texts was conducted by Balahur and Steinberger [2]. The researchers tried to define the scope of the task; separating good from bad news content. They also tried to analyze clearly marked opinions that were explicitly expressed in news texts.

Tony Mullen and Robert Malouf [27] carried out statistical tests on political discussion group postings. This work was a preliminary work in the sentiment analysis of Informal Political Discourse.

Our work is new since no previous research has been conducted in the sentiment analysis of news texts for Turkish and there are just a few initial studies conducted in languages other than English. Besides, due to the lack of research, there are no results indicating how well different machine learning methods could perform in Turkish sentiment analysis.

## 2.3 Sentiment Analysis with Transfer Learning

There are lots of sentiment analysis techniques and different areas that these tech-niques are applied. Transfer learning and domain adaptation techniques are used widely in ML [38]. Transfer learning has been applied in many different research areas containing NLP problems, learning data across domains, image classification problems etc. [29]

One of the problems that transfer learning and domain adaptation are applied is sentiment classification. Ave and Gommon conducted an initial study to customize sentiment classifiers to new domains [1]. Bliter et al. extend Structural Correspondence Learning (SCL) to sentiment classification to investigate domain adaptation for sentiment classifiers by focusing on

online reviews for different types of products [21]. Li et al. outline a novel sentiment transfer mechanism based on constructed non-negative matrix tri-factorizations of term document matrices in the source and target domains [38].

In our work, different than the other domain adaptation and transfer learning methods applied in sentiment classification tasks, we use unlabeled data with unsupervised feature construction, and transferring knowledge from short text (tweets) to long text (columns), which is not a common technique applied in transfer learning. Besides, our work is an initial work for applying transfer learning for sentiment classification of Turkish texts.

## 2.4 Sentiment Analysis in Turkish

Sentiment classification of Turkish texts has become an area of interest for researches lately, and there is still not much work in the field. One of the first studies in Turkish is Erogul's master thesis work [15], which is Sentiment Analysis in Turkish. By using collected Turkish movie review data, he tried to classify reviews by using SVM with some NLP techniques. As features with n-grams he also examined the effects of part of speech tagging, spellchecking and stemming. He succeeded about 85% accuracy.

Boynukalın in her master thesis (2012) worked on Emotion Analysis of Turkish texts [6]. In her study, she has introduced an emotion classification study on Turkish texts. A new dataset for Turkish emotion analysis is introduced. By trying several classification techniques, she has compared the results. Besides, due to the morphological characteristics of Turkish language she has added new features to improve the performance.

Cakmak et al. (2012) describe a methodology showing the feasibility of a fuzzy logic representation of Turkish emotion related words. They conclude that there is strong correlation between emotions attributed to Turkish word roots and the Turkish sentences [8].

Vural et al. (2013) present a framework for unsupervised sentiment analysis in Turkish documents. They test with Turkish movie reviews and although they use unsupervised approach, their framework performs nearly as successful as supervised techniques [42].

In their work, Özsert and Özgür (2013) tried to determine the polarity of words using a multilingual approach. They basically construct a graph (word relatedness) for multi languages and tried to produce set of positive and negative seed words by using English seed words [46].

A shorter version of our method on our work on sentiment analysis appeared in [22]. Our approach including transfer learning appeared in [23].

# CHAPTER 3

# BACKGROUND

In this chapter, the background information about technologies, methods and some evaluation metrics are given. Background information about different machine learning methods used are given under machine learning part. Besides, background information about NLP, transfer learning are given. The definitions of TF-IDF weighting and evaluation metrics used in the experiments are also given in this chapter.

## 3.1 Machine Learning

Machine learning (ML) which is a subfield of artificial intelligence (AI) refers to recognize, plan, control, predict information from data. To extract and learn information automatically from raw data, ML uses statistical and computational methods.

In this thesis to examine the sentiment classification of columnists' columns, we use four different ML algorithms: the Support Vector Machine, the Naive Bayes Classification, Maximum Entropy Classification and the N-gram based character Language Model.

We explain each of the ML methods in detail.

### 3.1.1 Naive Bayes Classification

This method is often used in text classification because of its simplicity and speed. Basically, Naïve Bayes makes the assumption that features (words) are generated independently of word position. It assigns a given document $d$ to the class:

$$c^* = argmaxcP(c|d)$$

We derive the Naïve Bayes Classifier as follows:

$$PNB(c|d) = P(c)\frac{\left(\prod_{i=1}^{m} P(f_i|c)^{n_i(d)}\right)}{P(d)}$$

where $f_i$'s are features that appear in the document and $n_i(d)$'s are the number of times features occur in the document.

### 3.1.2 Support Vector Machines

The SVMLight[1]is used for training and testing purposes by using the default parameters of the package. The idea behind SVM is as follows: in the training phase to find a hyper plane separating the document vectors in one class from those in the other and making the margin or separation as large as possible. In the testing phase the aim is to classify test instances based on which side of the hyperplane they fall on. Let us say that the hyperplane is represented as follows:

$$w = \sum_j \alpha_j c_j d_j, \quad \alpha_j \geq 0$$

$w$ is a vector that represents the hyperplane. $cj \in \{-1, 1\}$ (-1 for negative, 1 for positive) is the correct calls of document $dj$. $\alpha j$'s are obtained by solving a dual optimization problem.

### 3.1.3 Maximum Entropy Classification

We use Apache OpenNLP[2]which is a machine learning based toolkit for the processing of natural language text that has maximum entropy implementation.

To assign a given document d to a class c. Maximum entropy uses the following exponential form:

$$P_M(c|d) = \left( \frac{1}{Z(d) exp(\sum_i \lambda_{i,c} F_{i,c}(d,c))} \right)$$

$Z(d)$ is a normalization function. $\lambda_{i,c}$'s are feature-weight parameters that are set so as to maximize the entropy of the induced distribution. The main goal is to choose the model making the fewest assumptions about the data while remaining consistent with it. Ten iterations of the improved iterative scaling algorithm [36] are performed for parameter training.

$$F_{i,c}(d,c') = \begin{cases} n_i(d) & \text{if } n_i(d) \geq 0 \text{ and } c = c' \\ 0 & \text{otherwise} \end{cases}$$

### 3.1.4 N-Gram based character language model

The N-gram based character language model is derived from the N-gram language models. Instead of words, this model takes characters as the basic unit in the algorithm instead of words [9].

The model provides $p(s)$ defined for strings $s \in \sum^*$ over an alphabet of characters $\sum$. For a character $c$ and string $s$ the chain rule is:

$$p(sc) = p(s) \times p(c|s)$$

[1]http://svmlight.joachims.org/

[2]http://incubator.apache.org/opennlp/

The N-gram Markovian assumption restricts the context to the previous $n-1$ characters, taking:

$$p(c_n|s_{c_1c_2\ldots c_{n-1}}) = p(c_n|c_1\ldots c_{n-1})$$

Therefore the maximum likelihood estimator for N-grams is:

$$p(cs) = \frac{C(sc)}{\sum_c C(sc)}$$

where $C(sc)$ is the number of times the sequence $sc$ observed in the training data and $\sum_c C(sc)$ is the number of single-character extensions of $sc$. In this research, we used the LingPipe[3] DynamicLMClassifier to test the effect of using the N-gram based character language model. It depends on a language model with a form of Witten-Bell smoothing [9].

## 3.2 Natural Language Processing

Natural Language Processing is a technology that deals with human language appearing in many sources like, web pages, social media, e-mails, newspaper articles for thousands of languages and their varieties.

There are a wide range of NLP applications from automatic question answering to email spam detection, from machine translation to spelling and grammar correction. Sentiment analysis aiming to extract opinion from a given text is an application of NLP, too.

Statistical NLP, dealing with machine learning and data mining methods, by using statistical, probabilistic and stochastic methods makes thousands or millions of possible analysis of the texts easier.

Combining machine learning techniques with NLP methods, it is possible to extract and analyze the opinion, sentiment of a given text from word level to sentence level and from the whole document to context level.

### 3.2.1 Natural Language Processing Tool (Zemberek)

To investigate the effects of stemming, spellchecking and part of speech tagging we used a Turkish stemmer developed under Zemberek [4] which is an open source natural language processing (NLP) framework for Turkic languages. In Zemberek, for stemming, a morphological parser basically finds the possible root and suffixes of a given word. We use the possible roots as features in the experiments. We use the token itself if the stemmer cannot find a possible stem.

[3] http://alias-i.com/lingpipe/

[4] http://code.google.com/p/zemberek/

Table3.1: The suggestions of the Zemberek library for the word “seçimlerinden”

| |
|---|
| sevimlerinden, seçiklerinden |
| seçimlerindin, seçimlerinden |
| seçimlerinsen, seçimlerinken |
| seçimlerinde, seçimlerimden |
| seçişlerinden, seçimlelinden |
| sezimlerinden, geçimlerinden |
| selimlerinden, seçilerinden |
| Serimlerinden, sehimlerinden |

### 3.2.2 Spellchecking

Spellcheck in NLP is to check the words in a text that may not be spelled correctly. We use Zemberek library for spellchecking purposes.

We use an extended spellchecking algorithm on input strings. The algorithm available under Zemberek allows us to handle up to 3 misplaced or wrong characters in the roots, 2 characters in suffixes. In Table 3.1 an example spellchecking can be seen for the word “seçimlerinden”, that can be translated as “from their choices”.

### 3.2.3 Part of Speech

The linguistic category that a lexical item belongs to is Part-of-Speech (POS). Adjective, noun and verb are common linguistic categories. Part of speech tagging is known as tokenizing a sentence and giving them most probable linguistic categories to the tokens.

We use Zemberek library for POS tagging purposes. We try to evaluate the performance of ML classifiers by providing POS information. We basically use three common linguistic categories in the experiments:

- noun
- adjective
- verb

## 3.3 Transfer Learning

Transfer Learning’s main goal is to extract useful knowledge from one or more source tasks and to transfer the extracted information into a target task where the roles of source and target tasks are not necessarily the same [29].

In our work, we aim to solve sentiment classification of Turkish political columns (target task) by extracting and transferring features from unlabeled Twitter data in an unsupervised way

(source task). Source domain is Twitter and contains unlabeled data; target domain is news and contains labeled data. Notice that source and target data does not share the class labels. Besides, the generative distribution of the labeled data is not the same as unlabeled data's distribution.

Our main motivation is the assumption that even unlabeled Twitter data collected from columnist's verified accounts may help us to learn important features in the politic news domain. By using this assumption, we use transfer learning. This kind of transfer learning is categorized as self-taught learning which is similar to inductive transfer learning [29]. Self-taught learning was first proposed by Raina et al. (2007) [31].

### 3.3.1 F-Score for Feature Ranking

In order to measure the importance of a feature for a classifier, we use F-Score (Fisher score) [11], [10]. F-Score has been chosen, since it is independent of the classifiers, so that we can use it for 3 different classifiers we use in experiments.

Given the training instances $x_i$ where $i = 1,2,3 \ldots l$ the F-score of the $j^{th}$ feature is defined as follows:

$$F(j) = \frac{(\bar{x}_j^{(+)} - \bar{x}_j)^2 + (\bar{x}_j^{(-)} - \bar{x}_j)^2}{\frac{1}{n_+ - 1}\sum_{i=1}^{n_+}(x_{i,j}^{(+)} - \bar{x}_j^{(+)})^2 + \frac{1}{n_- - 1}\sum_{i=1}^{n_-}(x_{i,j}^{(-)} - \bar{x}_j^{(-)})^2}$$

where $n_+$ and $n_-$ are the number of positive and negative instances in the data set respectively; $\bar{x}_j$, $\bar{x}_j^+$ and $\bar{x}_j^-$ represents the averages of the $j^{th}$ feature of the whole positive-labeled and negative-labeled instances; $\bar{x}_{i,j}^+$ and $\bar{x}_{i,j}^-$ represent the $j^{th}$ feature of $i^{th}$ positive and negative instance. Larger F-Score means that the feature has more importance for the classifier.

Another method to determine the importance of each feature is to check how the performance has been affected without the existence of that feature by simply extracting each feature from the training data and evaluating the classifier again [10]. Usually this method is very expensive when the number of features is very large.

## 3.4 TF-IDF Weighting

Term frequency-inverse document frequency measures how important a feature (word) to a document and it is used as weighting factor in text mining applications. Variations of tf-idf calculations are available, and in this work we use the following formulations [12]:

Given a corpus $D$, a document $d$ and a term $t$ in that document term frequency, inverse term frequency and tf-idf are calculated by multiplying tf and idf:

$$tf = \frac{\text{number of times } t \text{ occurs in } d}{\text{total number of terms in } d}$$

Table3.2: Contingency table for performance evaluations

| | Actual Class(Expectation) | |
|---|---|---|
| Predicted Class(Observation) | tp(true positive)<br>tn(true negative) | fp(false positive)<br>fn(false negative) |

$$idf = log\left(\frac{\text{number of docs in } D}{\text{number of docs in } D \text{ that } t \text{ occurs in}}\right)$$

## 3.5 Evaluation Metrics

To evaluate the performance of the sentiment classification of news data with four different Machine Learning methods, we adopted Accuracy, Precision, Recall and F1-Measure metrics that are generally used in text categorization and information retrieval. The values in Table 3.2 are as follows:

- True positive = correctly classified as positive
- False positive = incorrectly classified as positive
- True negative = correctly classified as negative
- False negative = incorrectly classifed as negative

These metric can be calculated according to the values in Table 3.2 and the following formulas:

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn}$$

$$\text{Precision} = \frac{tp}{tp+fp}$$

$$\text{Recall} = \frac{tp}{tp+fn}$$

$$\text{F1-measure} = 2 \times (precision \times recall)/(precision + recall)$$

The degree of closeness of a measured quantity to its actual value is called as Accuracy. In this work, Accuracy is the number of columns classified correctly by classifiers, divided by the total number of columns tested.

The proportion of positive test results which are true positives is known as Precision.

The proportion of actual positive documents that are correctly classified as such is known as Recall.

F1-measure combines Precision and Recall by calculating the weighted harmonic mean of both.

# CHAPTER 4

# PROPOSED METHODS/APPROACHES

In this chapter, we present the details of the proposed methods. In Section 4.1 the detailed description of the data set used in this research is given. In Section 4.2 the details of the methods used by using the explained data sets are given.

## 4.1 Data Description

In this research, the work is based on datasets from two different domains:

- News
- Twitter

For News Domain, there are two data sets that we created:

- Turkish Political Columns of Turkish Columnists
- English Political Columns of Turkish Columnists.

For Twitter Domain, there are also two data sets: Tweets from columnists' Twitter accounts and tweets from random Twitter accounts. In Table 4.1 the details about the used data sets are given.

Table4.1: Data sets used

| Data Set | Size | Type |
|---|---|---|
| Turkish News | 400 columns | Labeled |
| English Columns | 50 columns | Labeled |
| Columnist Twitter Account | 123074 tweets | Unlabeled |
| Random Tweets | Over 100000 tweets from random accounts | Unlabeled |

Table4.2: Sample from Turkish Political News Column

| Textual Details of News Data | Label |
|---|---|
| Türkiye'de bugünlerde Irak petrolünden çocuk gelinlere hemen hemen bütün konuların bir şekilde Kürt problemiyle ilişkilendirilmesi belki de Türkiye toplumunda ilk kez çözüm odaklı bir isteğin olduğunu ve bu şekilde ilerlenmesi gerektiğini gösteriyor. (...It is interesting that almost every issue in Turkey nowadays, from Iraqi oil to child brides, is somehow linked to the Kurdish problem, which is one indication that there is a will, for perhaps the first time in Turkish society, to get the problem solved and move on...) | pos |

### 4.1.1 Turkish Political Columns of Turkish Columnists

Our data came from 51 differet columnists of 6 different Turkish newspapers. We collected more than 1000 columns and after selection and annotation processes we had a total of 400 columns, 200 positive and 200 negative. The collected data was annotated by three native speakers of Turkish.

In the selection process annotaters tried to look at columns that are related to politics domain. For instance, if a column was containing textual data about both sports and politics, that column is eliminated. The 400 collected and annotated columns are the ones for which the three annotators agreed about whether they showed positive or negative criticism, or none. While annotating the columns as support or critism, annotaters focused on the overall content of the columns. For instance, if the columnist wrote about politics of Middle East and claimed that a possible war would result in death of thousands of civilians, that specific column is annotated as critisim since the columnist is explaining his opinions about the damages of the war.

During the data collection period we observed that most of the time columnists express criticism rather than support. Therefore, finding documents containing positive criticism was a difficult task. Another observation was that columnists sometimes write about 2-3 unrelated topics in the same column. Since we were looking at the overall sentiment of the column; we did not use such columns in our data set.

The formulation of labeled news data that will be used in the rest of the paper is as follows:

$$T = \left\{ (x_l^{(1)}, y^{(1)}), (x_l^{(2)}, y^{(2)}) \ldots, (x_l^{(m)}, y^{(m)}) \right\}$$

A news data, such as the one shown in Table 4.2, is represented as $(x_l^{(j)}, y^{(j)})$ where $x_l^{(j)} = (f_1, f_2 \ldots, f_k)$ is a term vector of the text and each $f_k$ is tf-idf (the definitions of tf and idf in are given in TF-IDF Weighting Section) value for features of the sample data and $y^{(j)} \in \{pos, neg\}$.

### 4.1.2 English Political Columns of Turkish Columnists

We collected some English columns from Hurriyet Daily News[1] and annotated 25 support and 25 criticism articles. The reason why we collected columns from this website is that, most of the English columns available there are written by Turkish columnists and they are political columns. Besides, the topics that are covered are similar to the Turkish political news data used in the previous experiments.

### 4.1.3 Tweets from columnists' Twitter accounts

Tweets from columnists' Twitter accounts are collected by using Twitter4J API[2]. Search API used to collect all accessible tweets of the columnists. Tweets of 51 Turkish columnist are collected and the number of tweets collected are 123074.

Notice that collected tweets are not annotated, this data set is used to construct features in an unsupervised way to be used in transfer learning.

The formulation of unlabeled Twitter data collected from columnist's Twitter account $U_1$ is as follows:

$$U_1 = \left\{ z_{u1}^1, z_{u1}^2 \ldots, z_{u1}^a \right\}$$

In $U_1$ each $z_{u1}^a$ corresponds to an unlabeled tweet of columnists' verified Twitter accounts.

### 4.1.4 Tweets from Random Twitter accounts

In order to collect random tweets Streaming API of Twitter4J is used. Over 100000 tweets are collected from random Twitter accounts.

Notice that collected tweets are not annotated, this data set is just used to eliminate noisy features from the data set, Tweets from columnists' Twitter accounts.

The formulation of unlabeled Twitter data collected from random Twitter accounts $U_2$ is as follows:

$$U_2 = \left\{ z_{u2}^1, z_{u2}^2 \ldots, z_{u2}^b \right\}$$

In $U_2$ each $z_{u2}^b$ corresponds to an unlabeled tweet of random Twitter accounts and they contain number of occurrences of each feature within tweet.

[1] http://www.hurriyetdailynews.com/
[2] http://twitter4j.org/en/index.html

## 4.2 Methods Used

We conducted K-fold-cross-validation, Kohavi 1995 [25], in the experiments, adopting K to be 3. Each time labeled news items are used to make a 3-fold cross validation in the data experiments. The data were partitioned randomly into three folds. On each round of experiments, we used 2-folds as the training data and the remaining fold as testing data.

To prepare the documents we removed all HTML tags automatically from the original document format . No stopwords removal was used since most of the stopwords carry strong sentiment.

Notice that punctuation marks are not eliminated; they are treated as separate lexical items.

For each experiment conducted with or without transfer learning the following steps are applied:

- For each news column in the training data a preprocessing stage is applied. At the beginning of the preprocessing stage each of the columns are split into sentences by using sentence detection algorithms. For sentence boundary detection we use LingPipe. After detecting the sentences we tokenize each detected sentence into words. Finally, we convert all the tokenized words' characters into lowercase.

- Notice that for news columns we do not use spellcheck method, since all of the columns are formal texts. However, for transfer learning purpose while extracting features to transfer we use spellcheck method to drop non-recognized words from tweets. It is important since tweets are informal texts and consist of many typo errors .

- If we use POS information in the experiments and if a certain part-of-speech is selected, the words that are not member of selected POS are eliminated .

- Preprocessing stage for unlabeled data (tweets) includes eliminating most frequent and less frequent features. Namely by using unlabeled data set $U_2$ we eliminated some features from unlabeled data set $U_1$.

- After preprocessing part, feature selection part is applied. In order to create features we use N-gram method. In most of the experiments we use unigrams as features. For some cases bigrams are considered also. Unigrams or bigrams are considered as features in the experiments. Right after creating the N-grams, each news column is converted to a real valued array that represents the features and their values. Values differ according to presence or frequency of the features. Presence vs. frequency problem is explained below in detail.

- Notice that feature selection part is different for the experiments that transfer learning is applied. The details are given in the algorithms in Section Transfer Learning .

### 4.2.1 Feature Selection

For feature selection we use the following bag-of-words framework:

Table4.3: Sample adjectives used in experiments

| Adjective | English meaning |
|---|---|
| acayip | odd |
| ayrı | different |
| beyaz | white |
| bol | loose |
| boş | empty |
| bütün | complete |
| canlı | alive |
| çıplak | naked |
| dağınık | scattered |
| dar | narrow |
| daracık | slinky |
| dolu | full |
| düz | smooth |
| geniş | wide |
| karışık | complicated |
| kısa | short |
| kolay | easy |
| kötü | bad |
| perişan | miserable |
| pis | foul |
| zayıf | weak |
| şirin | cute |

Let $\{f_1, f_2 \dots f_m\}$ be a predefined set of m features that can appear in a document $d$. Let $n_i(d)$ be the number of times $f_i$ occurs in a document $d$, then each document is represented as follows: $d = (n_1(d), n_2(d) \dots n_m(d))$. For N-Gram, the technique used is explained in section "N-gram based character language model".

To analyze the effects of different features selected, we conducted experiments by using different selected features:

- Unigram: Unigrams are single words occurring in training and test data documents. For example, "demokrasi (democracy)", "yanıltıcı (misleading)" are example unigrams used in the experiments.
- Unigram + Adjectives: Adjectives are treated as separate features and used as additional features with unigrams. In Table 4.3 sample adjectives are given.
- Unigram + Effective Words: Effective words shown in Table 4.4 are treated as separate features and used with unigrams.
- Bigram: Contiguous sequence of 2 words. For example, "adil değil (not fair)", "olumlu şekilde (positively)".

Table4.4: Turkish effective words, their english meanings

| Proposed Effective Words | |
|---|---|
| Positive | Negative |
| cesaret(courage) | yasak(forbidden) |
| teşekkür(appreciation) | yüzsüz (impudent) |
| minnet (gratitude) | arsız (sassy) |
| gönülden (lief) | vahim (desperate) |
| emek (labour) | utanç (shame) |
| mümkün (possible) | rezil (outrageous) |
| destek (support) | kötümserlik (pessimism) |

### 4.2.2 Using Root of Words

To investigate the effects of stemming we used a Turkish stemmer developed under Zemberek which is explained in Background Section .

### 4.2.3 Presence vs. Frequency

In Information Retrieval, text is represented as feature vector that contains the frequencies for individual terms within the text. Since Pang and Lee (2002) [30] obtained better performance results for SVM by using presence of the features we use presence information in experiments as a method also.

In the bag-of-words explained in the previous section a document is shown as follows: $d = (n_1(d), n_2(d) \ldots n_m(d))$ for m different features. For presence we simply convert $n_i(d)$ to 1 if it is greater than 0, otherwise it remains as 0.

Notice here that for the N-gram character based language model the usage of different features and frequency, and the presence of the features are different than for the 3 other algorithms. The steps followed are as follows:

- While creating the language models, input text is tokenized. For presence each different token is used only once, for frequency all the tokens are used.
- For stemmed unigrams, the stems of the tokens are given to language models as input data.
- For adjectives and unigrams, each token is identified as adjective, effective word or not. If they are adjectives or effective words the token is given to the language model twice.

Pang and Lee (2008) [7], observed that binary-valued features (presence information) brings more useful information for classification of review-related texts. Notice that, in this work we try to classify columns, which are much longer than reviews, and using presence information may not result with better performance.

### 4.2.4 Transfer Learning Approach

We use variants of the algorithm proposed by Raina et al. (2007) [31] for self-taught transfer learning which is known as inductive transfer learning [29] in our problem. Raina et al. basically consider to solve a supervised task given labeled data and unlabeled data by constructing features from unlabeled data in an unsupervised way.

We propose two different approaches in the unsupervised construction of the transferred features. Some formulations used in the algorithms are given below:

$$U_1 = \left\{ z_{u1}^1, z_{u1}^2 \ldots, z_{u1}^a \right\}$$

$$U_2 = \left\{ z_{u2}^1, z_{u2}^2 \ldots, z_{u2}^b \right\}$$

$$L_u = \left\{ (f_u^1, v^1), (f_u^2, v^2) \ldots, (f_u^a, v^a) \right\}$$

$$T = \left\{ (x_l^{(1)}, y^{(1)}), (x_l^{(2)}, y^{(2)}) \ldots, (x_l^{(m)}, y^{(m)}) \right\}$$

$$L_v = \left\{ (f_v^{(1)}, F(f_v^{(1)})), (f_v^{(2)}, F(f_v^{(2)})) \ldots, (f_v^{(a)}, F(f_v^{(a)})) \right\}$$

Notice that, in both of the algorithms explained in detail below, by using less frequent and most frequent features in $U_2$, which is a list of features (unigrams) collected from random twitter accounts, noisy features are eliminated from $U_1$, which is a list of features collected from columnists' twitter accounts. Then, by using filtered $U_1$, a list $L_u$ of sorted features according to their occurrences in the all documents is generated. Then, the number of occurrences of the features are normalized using the $log_x$ function (best x is chosen after several experiments). Actually the normalized list $L_u$ contains feature and value pairs. Notice that, after normalization some features in $L_u$ are eliminated. For the second algorithm described below, by using the labeled training set $T$, we calculate F-score of each feature and a list $L_v$ of sorted features according to their F-scores is generated.

**Algorithm 1:** Unsupervised feature construction without feature rankings

**Input**: $T$, $U_1$ and $U_2$
**Output**: Learned Classifier $C$ for Classification Task

1 Filter $U_1$ by using less and most frequent features in $U_2$
2 Construct $L_u$ by using $U_1$
3 By using $T$ and $L_u$ construct new labeled set $\bar{T} = \left\{ (\bar{x}_l^{(i)}, y^{(i)}) \right\}_{i=1}^{m}$
4 Learn a classifier $C$ by applying supervised learning algorithm (SVM, Naive Bayes or Maximum Entropy).
5 **return** $C$

In Algorithm 1 unsupervised feature construction without using the knowledge of the feature rankings within the classifier used.

Without transfer learning tf-idf values for $T$ are as follows:

$$\forall i \forall j \left( x_l^{(i)}(f_j) \right) = \frac{count\left( (x_l^{(i)}(f_j)) \right)}{j} \times idf$$

After applying steps 1, 2 and 3 in Algorithm 1, with transfer learning (we simply increase the term frequency of transferred features) the tf-idf in $\bar{T}$ become as follows:

$$\forall i \forall j \left( x_l^{(i)}(f_j) \right) = \frac{count\left( (x_l^{(i)}(f_j)) \right) + log_x(valueoff_j inL_u)}{j} \times idf$$

In the Algorithm 1 explained above, we have performed two different experiments. In the first set of experiments while constructing $\bar{T}$ only the transferred features of $L_u$, are included into $\bar{T}$. Namely, features in the target domain that do not appear in $L_u$ are eliminated. In the second set of experiments conducted by following steps in Algorithm 1, $\bar{T}$ contains both the transferred features from $L_u$ and the raw features of $T$. That is, only step 3 differs.

**Algorithm 2:** Unsupervised feature construction with feature rankings

**Input**: $T$, $U_1$ and $U_2$
**Output**: Learned Classifier $C$ for Classification Task

1 Filter $U_1$ by using less and most frequent features in $U_2$
2 Construct $L_u$ by using $U_1$
3 Use $T$ to calculate f-score of each feature and $L_v$.
4 Combine $L_u$ and $L_v$ and have a list $L_{u+v}$.
5 Transfer knowledge in $L_{u+v}$ to $T$ and obtain $\bar{T} = \left\{ (\bar{x}_l^{(i)}, y^{(i)}) \right\}_{i=1}^{m}$
6 Learn a classifier $C$ by applying supervised learning algorithm (SVM, Naive Bayes or Maximum Entropy).
7 **return** $C$

In Algorithm 2 unsupervised feature construction with using the knowledge of feature rankings within the classifier used.

By combining $L_u$ and $L_v$ a third list $L_{u+v}$ is constructed as follows:

$$L_{u+v} = c_1 L_u + c_2 L_v$$

In order to decide on optimal $c_1$ and $c_2$ values several experiments are conducted. Different than Algorithm 1, after transferring information the tf-idf in $\bar{T}$ become as follows:

$$\forall i \forall j \left( x_l^{(i)}(f_j) \right) = \frac{count\left( (x_l^{(i)}(f_j)) \right) + log_x(valueoff_j inL_{u+v})}{j} \times idf$$

Similar to Algorithm 1, in Algorithm 2 two different experiments are performed. $\bar{T}$ is constructed by using only the transferred features from $L_u$, by using both transferred features from $L_u$ and raw features of $T$.

To summarize the algorithms, we propose two different transfer learning algorithms: In the first algorithm, we increase the term frequency of a feature within the unlabeled data by using its total number of occurrences (by normalization) and in the second one we similarly increase the term frequency of a feature by using unlabeled data and the F-score of the feature within labeled data.

# CHAPTER 5

# RESULTS AND DISCUSSION

In this chapter, the results of the several experiments conducted are given, analyzed and discussed in detail.

## 5.1 Results For Experiments without Transfer Learning

### 5.1.1 Baseline

For creating a baseline we use the claim (Pang and Lee, 2002) [30] that there are certain words people tend to use to express strong sentiments, so that it might suffice to simply produce a list of such words to classify texts. We chose good indicators for positive and negative columns, 197 positive and 300 negative indicators.

Some of our selections are shown in Table 5.1. We converted these words into simple decision procedures counting the number of the proposed positive and negative words in a given document. The accuracy for a simple baseline classifier technique gives 59% accuracy which might be considered as a baseline. We used this preliminary experiment as baseline for other experiments.

Table5.1: Turkish indicators, their english meanings

| Proposed Indicators | |
|---|---|
| Positive | Negative |
| cesaret(courage) | yasak(forbidden) |
| teşekkür(appreciation) | yüzsüz (impudent) |
| minnet (gratitude) | arsız (sassy) |
| gönülden (lief) | vahim (desperate) |
| emek (labour) | utanç (shame) |

Table5.2: Accuracy values by using different N in N-Gram Language Model

| N | Accuracy |
|---|---|
| 1 | 58.31 |
| 2 | 62.34 |
| 3 | 72.28 |
| 4 | 73.70 |
| 5 | 74.17 |
| 6 | 74.40 |
| 7 | 76.31 |
| 8 | **76.54** |
| 9 | 76.30 |
| 10 | 75.59 |
| 11 | 75.74 |

### 5.1.2 Selection of N for N-Gram based character language model

Note that we took N=8 in the case of the N-Gram model. We tested this using unigrams as features to compare different values of N. The results can be found in Table 5.2. We can observe that by taking N=8 we obtained the best performance. It can be seen that after 7 it makes no difference in performance. According to the results presented in Table 5.2 we took N=8 in the remaining experiments.

### 5.1.3 Overall Results

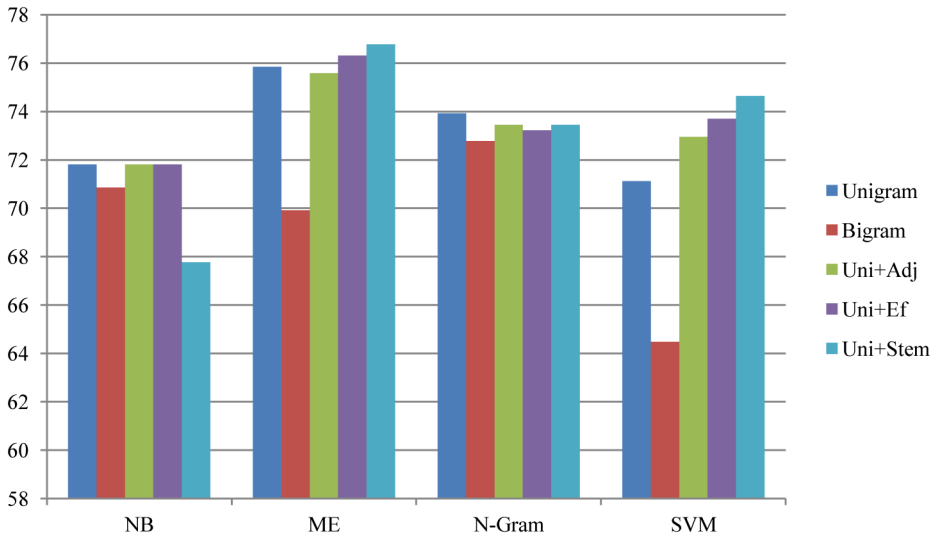


Figure 5.1: Accuracies for frequency experiments

Table5.3: Accuracies in percentage

| | Features | freq. or pres. | NB | ME | N-Gram | SVM |
|---|---|---|---|---|---|---|
| (1) | unigram | frequency | 71.81 | **75.85** | 73.93 | 71.12 |
| (2) | unigram | presence | 72.05 | 74.88 | **76.54** | 74.88 |
| (3) | bigram | frequency | 70.86 | 69.92 | **72.77** | 64.48 |
| (4) | bigram | presence | **72.05** | 69.44 | 71.80 | 66.81 |
| (5) | unigram+adjective | frequency | 71.81 | **75.59** | 73.45 | 72.95 |
| (6) | unigram+adjective | presence | 72.29 | 74.88 | 76.30 | **76.31** |
| (7) | unigram+effective | frequency | 71.81 | **76.31** | 73.22 | 73.70 |
| (8) | unigram+effective | presence | 72.05 | 76.06 | **76.78** | 74.65 |
| (9) | unigram(stemmed) | frequency | 67.77 | **76.78** | 73.45 | 74.65 |
| (10) | unigram(stemmed) | presence | 67.99 | 74.88 | 75.35 | **76.31** |

Table5.4: Precision in percentage

| | Features | freq. or pres. | NB | ME | N-Gram | SVM |
|---|---|---|---|---|---|---|
| (1) | unigram | frequency | 70.04 | 76.31 | 73.02 | **76.90** |
| (2) | unigram | presence | 70.17 | 75.44 | **75.93** | 72.81 |
| (3) | bigram | frequency | 69.11 | 77.50 | 71.98 | **77.65** |
| (4) | bigram | presence | 70.05 | **74.91** | 72.92 | 64.96 |
| (5) | unigram+adjective | frequency | 70.04 | **76.12** | 72.61 | 74.92 |
| (6) | unigram+adjective | presence | 70.06 | 75.23 | 75.46 | **75.64** |
| (7) | unigram+effective | frequency | 70.16 | **76.26** | 72.18 | 75.93 |
| (8) | unigram+effective | presence | 70.08 | **77.08** | 76.02 | 72.15 |
| (9) | unigram(stemmed) | frequency | 66.73 | 76.34 | 71.24 | **76.49** |
| (10) | unigram(stemmed) | presence | 62.85 | 74.53 | **74.81** | 73.63 |

Table5.5: Recall in percentage

| | Features | freq. or pres. | NB | ME | N-Gram | SVM |
|---|---|---|---|---|---|---|
| (1) | unigram | frequency | 76.79 | 74.92 | **77.29** | 60.24 |
| (2) | unigram | presence | 77.67 | 73.95 | 78.22 | **80.13** |
| (3) | bigram | frequency | **75.85** | 56.42 | 75.84 | 40.34 |
| (4) | bigram | presence | **77.77** | 58.81 | 71.12 | 73.95 |
| (5) | unigram+adjective | frequency | 76.79 | **77.75** | 76.78 | 69.25 |
| (6) | unigram+adjective | presence | 78.25 | 74.43 | **78.69** | 78.24 |
| (7) | unigram+effective | frequency | 76.32 | **76.82** | 76.71 | 70.16 |
| (8) | unigram+effective | presence | 77.27 | 77.08 | 78.69 | **81.07** |
| (9) | unigram(stemmed) | frequency | 72.04 | 77.74 | **80.10** | 71.09 |
| (10) | unigram(stemmed) | presence | 73.93 | 75.83 | 77.26 | **82.01** |

Table5.6: F1-Measure

| | Features | freq. or pres. | NB | ME | N-Gram | SVM |
|---|---|---|---|---|---|---|
| (1) | unigram | frequency | 72.76 | 72.64 | **75.09** | 67.56 |
| (2) | unigram | presence | 73.73 | 74.69 | **77.06** | 76.29 |
| (3) | bigram | frequency | 72.32 | 65.3 | **73.86** | 53.09 |
| (4) | bigram | presence | **73.71** | 65.89 | 72.01 | 69.16 |
| (5) | unigram+adjective | frequency | 73.31 | **76.93** | 74.64 | 71.97 |
| (6) | unigram+adjective | presence | 73.93 | 74.83 | **77.04** | 76.92 |
| (7) | unigram+effective | frequency | 73.11 | **76.54** | 74.38 | 72.93 |
| (8) | unigram+effective | presence | 73.5 | 77.08 | **77.33** | 76.35 |
| (9) | unigram(stemmed) | frequency | 69.28 | **77.03** | 75.41 | 73.69 |
| (10) | unigram(stemmed) | presence | 67.94 | 75.17 | 76.02 | **77.6** |

Figure 5.1 shows the accuracy values of four different methods by using frequency of the features. We can see that Maximum Entropy performs better than other algorithms.

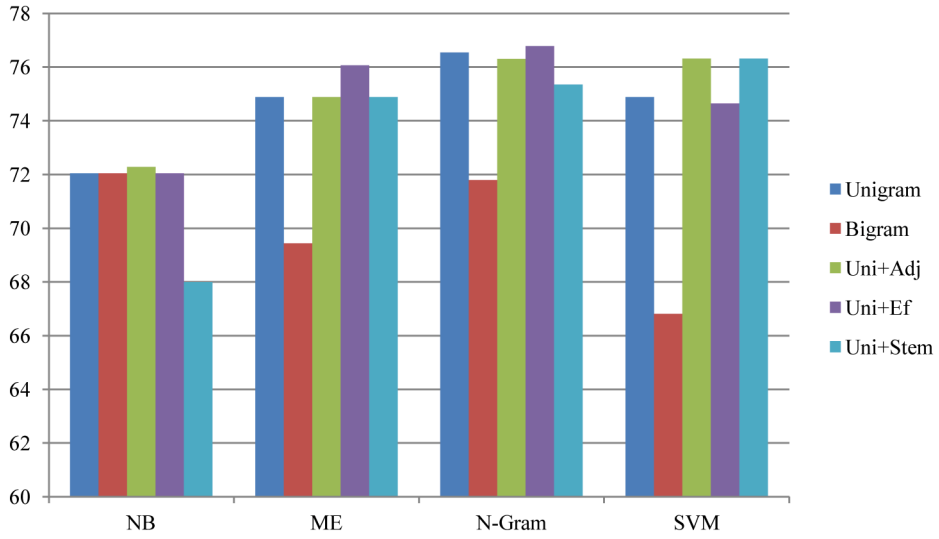


Figure 5.2: Accuracies for presence experiments

The overall accuracies of four algorithms using different features are indicated in Table 5.3. Table 5.4 shows the precision values of the experiments and in Table 5.5 we show the Recall values. Table 5.6 shows F1-Measure values for experiments. Figure 5.1 and Figure 5.2 indicate the graphics of accuracy values for frequency and presence information.

#### 5.1.3.1 Feature frequency vs. presence

In order to investigate the effect of the frequency and presence information, we ran all the experiments twice, once with the presence of the features only and once with the frequencies of the features only.

As can be seen from the results, frequency information does not have a positive effect on the sentiment classification of the columns except for the Maximum Entropy Classification. In almost all of the experiments adopted with other algorithms, better performance was achieved by accounting feature presence not feature frequency. Although the difference is not significant in the case of algorithms other than Maximum Entropy, presence information provides better results.

Figure 5.2 shows the accuracy values of four different methods by using presence of the features. We can see that N-Gram language model performs better than the other algorithms in most of the cases.

#### 5.1.3.2 Results of using different features

The classification accuracies resulting from using different features show that all of the experiments clearly surpass the baseline of 59%.

In the experiments, in addition to unigrams, we also studied the use of bigrams as features. Lines 3 and 4 of Table 5.3 show that bigram information does not improve performance beyond unigram information. We can see a 5% to 8% decrease for SVM, Maximum Entropy and the N-gram Language Model. In the case of Naïve Bayes there is not much difference between the results from using unigrams and bigrams. We can conclude that bigram information is not as useful as unigram information for the Sentiment Classification of news columns.

We also experimented effect of adjectives and effective words. To investigate the effects of adjectives and effective words carrying strong sentiment, unigram + adjectives and unigram + effective words were used as features.

One might expect that adjectives or effective words carry a great deal of information about the sentiment of a news column. The results indicate, however, that neither adjectives nor effective words provide a notably better performance.

We can see only a very small percentage improvement in the accuracies for different algorithms used, less than 1%. This may be due to the length of the documents used in the experiments. Unlike short movie reviews, political news columns are quite long. The frequency of adjectives and effective words used as features may be very small in the documents.

Using the stems of the unigrams as features makes no significant difference in the results either.

Comparing four different algorithms used in the experiments it can be observed that the N-gram character based Language Model outperforms other algorithms in the case of unigrams, at over 70%. Maximum Entropy is the second algorithm that performs well, although there is small difference between Naïve Bayes and Maximum Entropy. An interesting observation is that SVM performs worse than other algorithms. It has been shown that for Sentiment

Table5.7: Accuracies for Turkish vs. English

| | NB | ME | N-gram | SVM |
|---|---|---|---|---|
| Turkish(Frequency) | 69.90 | 64.12 | **74.07** | 69.68 |
| English(Frequency) | 67.59 | 75.69 | **79.66** | 73.61 |
| Turkish(Presence) | 63.66 | 65.97 | **69.90** | 67.82 |
| English(Presence) | 71.53 | 60.18 | **73.83** | 69.91 |

Classification of short reviews SVM performs better than Naive Bayes and Maximum Entropy [30].

Comparing the 4 algorithms according to their runtime performance we observed that SVM and Naïve Bayes perform better than Maximum Entropy and the N-Gram character based Language Model. However, as we can see from the accuracy results, Naïve Bayes performs worse than the other algorithms, and SVM in most of the tests cannot perform as well as Maximum Entropy and the Language Model.

#### 5.1.3.3 Turkish vs. English

In order to make a comparison between English and Turkish we made a further experiment. We collected some English columns from Hurriyet Daily News [1] and annotated 25 support and 25 criticism articles. The reason why we collected columns from this website is that, most of the English columns available there are written by Turkish columnists and they are political columns. Besides, the topics that are covered are similar to the Turkish political news data used in the previous experiments.

To make a coherent comparison we choose the same amount of Turkish data by using random sampling, 25 support and 25 criticism columns among 200 support and 200 criticism Turkish columns. We just tested for using unigram features for both frequency and the presence of the features. In all experiments 3-fold cross validation is used. The results are indicated in Table 5.7.

It can be observed that in most of the experiments we obtain better accuracies for English. Throughout 8 experiments in only 2 of them accuracy values for Turkish are better, Naïve Bayes with frequency feature by 2% and Maximum Entropy with presence feature by 5%. In the remaining 6 experiments accuracy values for English data outperforms accuracy values for Turkish data by 2% to 11%. The reason may be the morphological differences between two languages. Turkish is a morphologically rich language that makes effective use of linguistic information such as the functional meaning of modal affixes, the semantic scope of negation terms like no, not, and the semantic classes of words. Therefore the usage of the frequency or presence features only may not be sufficient for Turkish data.

[1] http://www.hurriyetdailynews.com/

Table5.8: POS Experiments

| | Method Used | Accuracy | Precision | Recall | F-Measure |
|---|---|---|---|---|---|
| POS(Noun) | NB | 69.45 | 67.29 | **76.29** | 71.39 |
| POS(Noun) | ME | 75.58 | 74.16 | **78.68** | 76.27 |
| POS(Noun) | N-Gram | 72.5 | 70.36 | **79.16** | 74.31 |
| POS(Noun) | SVM | 74.88 | 73.37 | **78.68** | 75.93 |
| POS(Adjective) | NB | 60 | 59.6 | **61.6** | 60.6 |
| POS(Adjective) | ME | 66 | **67.1** | 63.4 | 65.2 |
| POS(Adjective) | N-Gram | 62.7 | 61.31 | **69.66** | 65.16 |
| POS(Adjective) | SVM | 65.18 | 65.24 | **65.41** | 65.32 |
| POS(Verb) | NB | 60.91 | 61.06 | **61.13** | 61.09 |
| POS(Verb) | ME | 63.5 | 63.85 | **64.45** | 63.99 |
| POS(Verb) | N-Gram | 69.9 | **71.15** | 66.83 | 68.92 |
| POS(Verb) | SVM | 63.5 | **64.97** | 60.68 | 62.75 |

Table5.9: Baseline Results

| | NB | ME | SVM |
|---|---|---|---|
| Unigram | 71.81 | **75.85** | 71.12 |
| Unigram+adjective | 71.81 | **75.59** | 72.95 |
| Unigram+effective words | 71.81 | **76.31** | 73.70 |

### 5.1.4 POS Results

To investigate the effect of POS experiments are conducted by using three different common linguistic categories. Namely, noun, adjective and verb used in the experiments. Notice that POS experiments are conducted by using unigrams as features.

The results are indicated in Table 5.8. It can be seen that using adjective and verb in the experiments does not improve the performance, but decreases. In the case of noun the performance of SVM increases compared to the result in Table 5.3. We can say that noun as a POS provides useful information. However, generally POS information provided to classifiers does not improve the performance.

## 5.2 Results For Experiments with Transfer Learning

### 5.2.1 Baseline

In order to have a baseline, sentiment classification of news columns are generated by using unigrams as features without transferring any knowledge from Twitter domain(for this purpose, we use the values from the previous experiments, Table 5.3).

### 5.2.2 Results of Transfer Learning Experiments

Several experiments are conducted. In the first set of experiments, unsupervised feature construction for transfer learning is applied without using the feature ranking knowledge. The amount of transferred features is from 1% to 100%. For each case, two different results are generated:

1. Learning the classifier $C$ by using only the transferred features, and then evaluating the performance.
2. Learning the classifier $C$ by using both raw features and the transferred features.

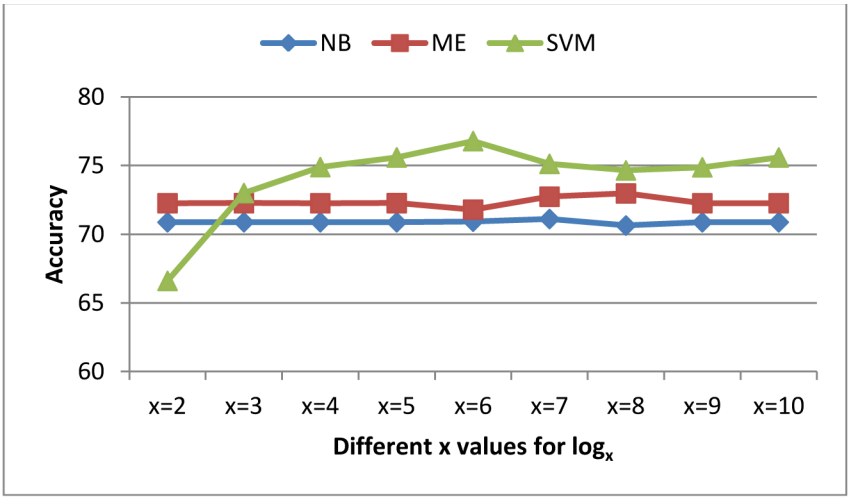


Figure 5.3: Accuracy values by using different $log_{(x)}$ in the first set of experiments

As explained in the previous section , we use $log_x(f)$ for normalization. Figure 5.3 indicates the results of classifiers for different x values. It can be seen that for x = 6 or x=7 almost the best results are obtained. Therefore, in the rest of the experiments in transfer learning part $log_6(f)$ is used to normalize the number of occurrences in the lists obtained from unlabeled data.

Figure 5.4 shows the performances of 3 different machine learning methods for varying amount of transferred features. Notice that these results are for the cases in which only the transferred features are included and no features from the target domain that is not in the source domain is included in the classification. In other words, while transferring the knowledge, for creating the classifier features, only those which are in the $L_u$ list are used. Features in the labeled data that are not in $L_u$ are eliminated. In this case SVM performed better than NB and ME. When compared with the baseline results, for SVM there is a 5.67% improvement. For NB there is small change and for ME there is a 4% decrease. Therefore, by using only the transferred features without feature ranking knowledge provides significant information only for SVM.

In Figure 5.5, the results are obtained by using the transferred features and the raw features together. Using only transferred features, gives better result for SVM and NB. Neither using

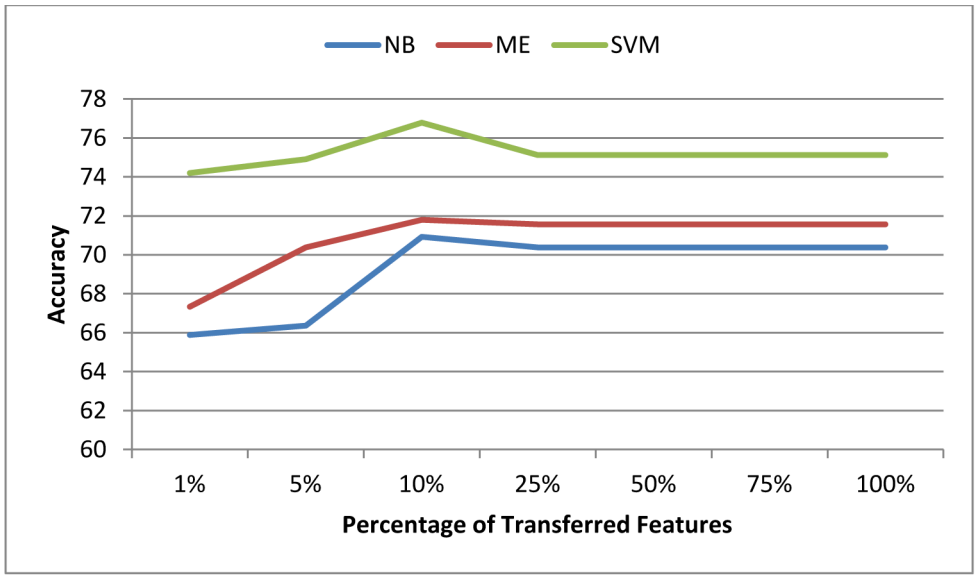


Figure 5.4: Accuracy values of Classifiers with only transferred features

only transferred features, nor using transferred features together, improves the performance for ME only.

In the second set of experiments, feature rankings obtained by using F-scores are used for unsupervised feature construction. In the previous section, the details of the method used are explained. In Algorithm 2 features are combined as $L_{u+v} = c_1 L_u + c_2 L_v$ list. In Figure 5.6, accuracy values for experiments are shown by using different $c_1$ and $c_2$ values. We observe that varying $c_1$ and $c_2$ values do not make significant changes. Therefore, we took $c_1 = 0.4$ and $c_2 = 0.6$ in the rest of the experiments.

Figure 5.7 shows the accuracy values when the amount of features transferred from constructed list $L_{u+v}$ varies. Combining two lists ($L_u$ obtained from unlabeled data and $L_v$ obtained from F-scores of the features) produces a very good performance gain. We can see from the Figure that especially transferring 5% of constructed $L_(u+v)$ list provides very useful information for the classification task for all techniques that we have tried. For NB up to 98.116% accuracy values are obtained. Comparing with the baseline results for NB, this corresponds to a 26.306% performance gain. We observe a 19.435% performance gain with a 93.135% accuracy value for ME. Finally, for SVM a 15.43% performance gain with a 91.74% accuracy value is reached. However, if the amount of the transferred features are in 10%-25% range of list $L_{u+v}$ , then the accuracy performance decreases for all techniques, and after that the results does not change. Roughly, we can say that features that carry important information for these classifiers are in the first 10% percentage of transferred features.

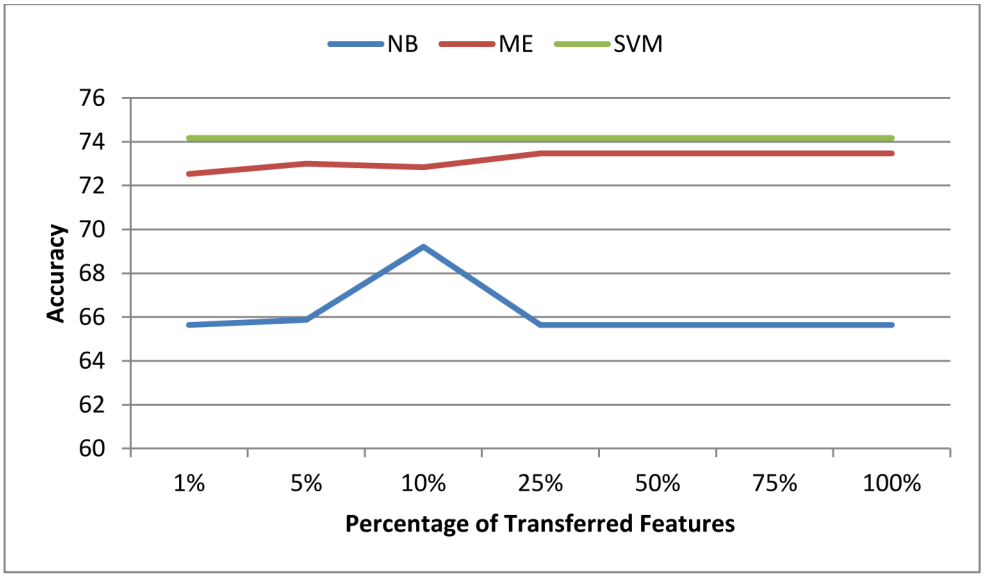


Figure 5.5: Accuracy values of Classifiers with the combination of transferred and raw features.

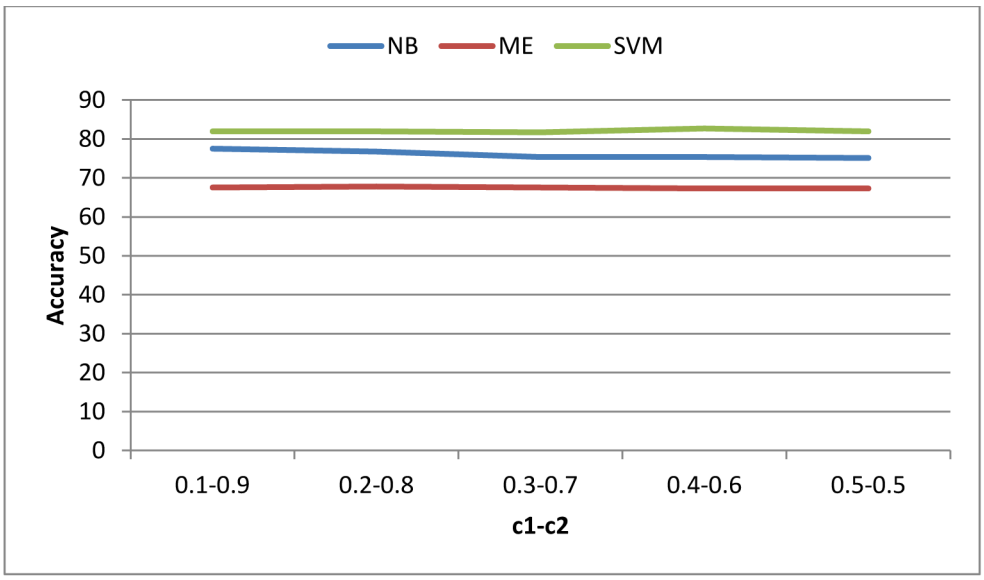


Figure 5.6: Accuracy values for f-score of features included for different values of c1-c2.

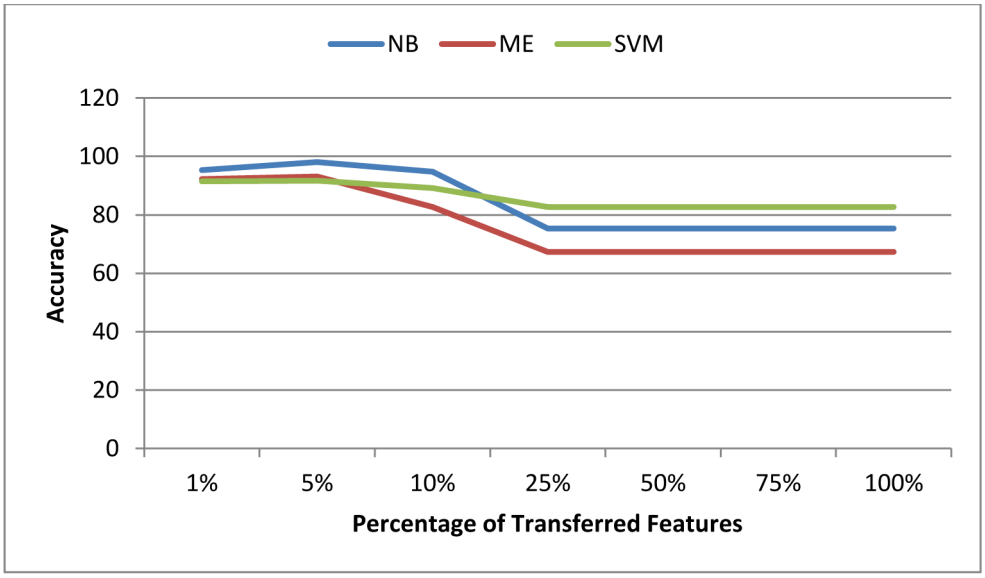


Figure 5.7: Accuracy values for f-score of features included for c1=0.4 and c2=0.6.

# CHAPTER 6

# CONCLUSION AND FUTURE WORK

## 6.1 Conclusion

Although the results produced via machine learning techniques in the experiments without Transfer Learning are quite good in comparison to the baseline results of 59% accuracy, our 76-77% accuracy is not good enough when compared to sentiment classification of reviews. For reviews, Pang et al. (2002) [30] and Boiy and Moens (2009) [5] obtained accuracies of up to 87.40% with SVM, Maximum Entropy and Naive Bayes. LingPipe for the N-Gram character based Language Model reported 81.5% accuracy for reviews by taking N=8. This can be explained by the difficulties explained in News Domain section. In terms of relative performance, the N-gram based character Language Model and Maximum Entropy performed better than Naïve Bayes and SVM. Unigram presences with the usage of adjectives turned out to be most effective for Naïve Bayes and SVM. Unigram presences with the usage of effective words were the most effective method for the N-gram Language Model. Unigram frequencies with the usage of stemming turned out to be most effective for Maximum Entropy. Although the usage of adjectives, effective words and stems provided better performance for different algorithms, the amount of improvement was not significant and so we cannot generalize their positive effect on sentiment classification in the news domain.

In the experiments with Transfer Learning, although we transfer knowledge from short text to larger text and transfer features from unlabeled data in an unsupervised way, this transfer learning method produced a very good improvement in the accuracy of the sentiment classification of Turkish political columns. Accuracy values over 90% for all three machine learning methods used (NB, SVM and ME) can be considered very important results for sentiment classification of news data. Similarly, performance improvement over 25% in transfer learning is also an important performance gain. In terms of relative performances, we see that in SVM, Naive Bayes and ME, transferred information improves the performance. It is also observed that, transferring features that are not in the first 10% of the transferred features decreases the performance.

We observe that the amount of transferred features do not make huge differences after a significant amount (25%). Besides, in the second set of experiments conducted by using F-score information of the features the best results are obtained by transferring 1% to 10% amount of features. This means that features carrying the most important information are the ones with higher frequency in the bag-of-words framework of transferred data.

An important outcome of our work is that, close accuracy values obtained by using raw fea-

tures and only transferred features shows that columnists' tweets are consistent with their columns. We can say that columnists use the same political knowledge and language both in the news domain and in the Twitter domain. Another important outcome is that, using feature ranking information (F score) combined with the unlabeled data turns out to be an effective method for transfer learning used in sentiment classification.

## 6.2 Future Work

For future work more training data and test data must be collected in order to generalize the results for Turkish news domain. Different experiments should be conducted to compare the results in a more convenient way.

After extensive experiments, the effects of the Turkish language may be analyzed and different methods may be used for news text classification for Turkish. For example, contextual shifters and subjectivity clues may be used. Morphological Analysis may be used also. Concerning the difficulties and challenges explained in Section 1.3 related to News Domain, data can be preprocessed by eliminating objective sentences with subjectivity-objectivity detection.

As future work, Named Entity Recognition might be used to detect which columnists write about which political party or politician and so on. Moreover, columnists can be clustered according to their sentiment (support or criticism) towards topics decided by NER.

Transferring from longer texts (different columns) can be analyzed, and using Transfer Learning in the Sentiment Classification of News Data with labeled data with Domain Adaptation techniques can be analyzed. Besides, using feature rankings together with unlabeled data can be adapted to different domains.

# REFERENCES


[1] A. Aue and M. Gamon. Customizing sentiment classifiers to new domains: a case study. *In http://research.micosoft.com/anthaue*, 2005.

[2] A. Balahur and R. Steinberger. Rethinking sentiment analysis in the news: from theory to practice and back. *WOMSA09*, pages 1–12, 2009.

[3] S. R. K. M. Z. V. G. E. . H. M. e. a. Balahur, A. Sentiment analysis in the news. *The 7th Conference on International Language Resources and Evaluation*, 2010.

[4] E. v. d. G. Balyaeva Evgenia. News bias of online headlines across languages. the study of conflict between russia and georgia. rhetorics of the media. *Conference Proceedings (2009) Lodz University Publishing House*, 2009.

[5] E. Boiy and M. Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Information Retrieval (2009)*, 12:526–558, 2009.

[6] Z. Boynukalin. Emotion analysis of turkish texts by using machine learning methods. *Middle East Technical University*, 2012.

[7] B.Pang and L.Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1):1–135, 2008.

[8] O. Cakmak, A. Kazemzadeh, S. Yildirim, and S. Narayanan. Using interval type-2 fuzzy logic to analyze turkish emotion words. *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, pages 1–4, 2012.

[9] B. Carpenter. Scaling high order character language models to gigabytes. *In Proceedings of the 2005 association for computational linguistics software workshop*, pages 1–14, 2005.

[10] Y. Chang and C. Lin. Feature ranking using linear svm. *In JLMR WCCI2008 workshop on casuality*, 3, June 2008.

[11] Y. Chen and C. Lin. Combining svms with various feature selection strategies. *In Isabelle Guyon, Steve Gunn, Masoud Nikravesh, and Lofti Zadeh, editors, Feature extraction, foundations and applications. Springer*, 2006.

[12] P. R. Christopher D. Manning and H. Schtze. Introduction to information retrieval. *Cambridge University Press, New York, NY, USA*, 2008.

[13] S. Das and M. Chen. Yahoo! for amazon. extracting market sentiment from stock message boards. alternation. *In Proc. of the 8th Asia Pacific Finance Association Annual Conference (APFA 2001)*, 2001.

[14] S. L. Dave, K. and D. M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *In Proceedings of WWW-2003*, 2003.

[15] U. Erogul. Sentiment analysis in turkish. master's thesis. *Middle East Technical University*, 2009.

[16] C. G. Fortuna Blaz and N. Cristianini. Detecting the bias in the media with statistical learning methods text mining. *Theory and applications Yatlor and Francis Publisher*, 2009.

[17] V. Hatzivassiloglou and K. McKeown. Predicting the semantic orientation of adjectives. *In Proc. of the 35th ACL/8th EACL*, pages 174–181, 1997.

[18] M. Hearst. Direction-based text interpretation as an information access refinement. *Text-based Intelligent Systems*, 1992.

[19] M. Hu and B. Liu. Mining and summarizing customer reviews. *In Proceedings of the KDD*, 2004.

[20] A. Huettner and P. Subasic. Fuzzy typing for document management. *In ACL 2000 Companion Volume: Tutorial Abstracts and Demonstration Notes*, pages 26–27, 2000.

[21] M. D. J. Blitzer and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. *Proc. 45th Ann. Meeting of the Assoc. Computational Linguistics*, pages 432–439, 2007.

[22] M. Kaya, G. Fidan, and I. Toroslu. Sentiment analysis of turkish political news. *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2012 IEEE/WIC/ACM International Conferences on*, 1:174–180, 2012.

[23] M. Kaya, G. Fidan, and I. Toroslu. Transfer learning using twitter data for improving sentiment classification of turkish political news. *to appear in ISCIS 2013*, 2013.

[24] S. M. Kim and E. Hovy. Determining the sentiment of opinions. *In Proceedings of the COLING*, 2004.

[25] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligenc*, 2(12):1137–1143, 1995.

[26] H.-J. C. Long-Sheng Chen, Cheng-Hsiang Liu. A neural network based approach for sentiment classification in the blogosphere. *Journal of Informetrics*, 5:313–322, 2011.

[27] M. R. Mullen, T. A preliminary investigation into sentiment analysis of informal political discourse. *In Proceedings of the AAAI symposium on computational approaches to analyzing weblogs*, pages 159–162, 2006.

[28] M. S. N. Godbole and S. Skiena. Large-scale sentiment analysis for news and blogs. *In Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

[29] S. J. Pan and Q. Yang. A survey on transfer learning. *Knowledge and Data Engineering, IEEE Transactions*, 22(10):1345–1359, October 2010.

[30] L. L. Pang, Bo and S. Vaithyanatham. Thumbs up? sentiment classification using machine learning techniques. *In Proceedings of the ACL-2002 conference on Empirical methods in natural language processing*, 2002.

[31] H. B. R. Raina, A. Battle and A.Y.Ng. Self-thaught learning: Transfer learning from unlabeled data. *Proc. 24th Int'l Conf. Machine Learning*, (759-766), June 2007.

[32] M. T. Rudy Prabowo. Sentiment analysis: A combined approach. *Journal of Informetrics*, 3:143–157, 2009.

[33] S. L. Rui Xia, Chengqing Zong. Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences*, 181:1138–1152, 2011.

[34] W. Sack. On the computation of point of view. *In Proc. of the Twelfth AAAI*, page 1488, 1994. Student Abstract.

[35] J. Z. Songbo Tan. An empirical study of sentiment analysis for chinese documents. *Expert Systems with Applications*, pages 2622–2629, 2008.

[36] V. D. P. Stephan Della Pietra and J. Lafferty. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):380–393, 1997.

[37] C. Strapparava and R. Mihalcea. Semeval 2007 task 14: Affective text. *In Proc. of ACL 2007*, 2007.

[38] C. D. Y. Z. Tao Li, Vikas Sindhwani. Bridging domains with words: Opinion analysis with matrix tri-factorizations. *In Proceedings of the Tenth SIAM Conference on Data Mining (SDM 2010)*, pages 293–302, 2010.

[39] R. M. Tong. An operational system for detecting and tracking opinions in online discussion. *SIGIR 2001 Workshop on Operational Text Classification*, 2001.

[40] P. Turney. Thums up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *In Proc. of ACL*, 2002.

[41] P. D. Turney and M. L. Litnemm. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *Technical Report EGB-1094, National Research Council Canada*, 2002.

[42] A. Vural, B. Cambazoglu, P. Senkul, and Z. Tokgoz. A framework for sentiment analysis in turkish: Application to polarity detection of movie reviews in turkish. *Computer and Information Sciences III*, pages 437–445, 2013.

[43] H. Yu and H. V. Towards answerring opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. *In Proceedings of EMNLP*, 32, 2003.

[44] W. H.-s. ZHU Jian, XU Chen. Sentiment classification using the theory of anns. *The Journal of China Universities of Posts and Telecommunications*, 17:58–62, July 2010.

[45] Z. Z.-Y. L. Ziqiong Zhang, Qiang Ye. Sentiment classification of internet restaurant reviews written in cantonese. *Expert Systems with Applications*, 2011.

[46] C. Özsert and A. Özgür. Word polarity detection using a multilingual approach. *Computational Linguistics and Intelligent Text Processing*, 7817:75–82, 2013.

# APPENDIX A

# EXAMPLE DATA

## A.1 Sample Turkish Political News Data

### A.1.1 Sample Positive Turkish Political News Data

Geçen hafta Dışişleri Bakanı Ali Babacan, bir "istişare toplantısı" düzenledi. Görüşmeye gazeteciler, akademisyenler ve sivil toplum kuruluşlarından davet edilen kişiler katıldı.

Önce Bakan Bey kısa bir konuşma yaptı; ardından katılımcılar görüşlerini beyan etti, bazen de sorular sordu. Babacan'ı dinlerken dikkatimi çeken en önemli noktalardan biri şuydu: Türkiye, dış politikada çok hızlı ve çok aktif bir rol üstlendi. O kadar ki çeyrek asra sığdırılan pek çok önemli gelişme, üç-beş senede yaşandı adeta. Bu durumun gelişen teknoloji ve hızlanan iletişim imkânlarıyla da ilgisi bulunmakta. Sebep ne olursa olsun gerçek şu ki, artık Türkiye "dört tarafı düşmanlarla çevrili ülke" olmaktan çıkıyor, herkesle belli bir diplomatik temas kuran, uluslararası problemlerde arabuluculuk yapan, sorun olmaktansa çözüme katkı sağlayan bir ülke durumundayız artık.

Dış politikada inisiyatif aldığımız hadiseleri bir çırpıda sayabilirsiniz. Mesela Afrika Zirvesi, İran'la ilgili arabuluculuk, Suriye ile İsrail'in bir araya getirilmesi, Şam'da düzenlenen dörtlü zirve, Kafkas Platformu için yapılan temaslar, millî maç vesilesiyle Ermenistan konusunda atılan adımlar... Bir yandan Cumhurbaşkanı, diğer yandan Başbakan, öbür taraftan Dışişleri Bakanı, teknokratlar, diplomatlar... Bütün bu temaslar arasında Türkiye'nin bir de Birleşmiş Milletler Güvenlik Konseyi'ne üyeliği için sarf edilen çabalar var. Babacan, kısa konuşmasının satır arasında, göreve geldiği andan itibaren 140 dışişleri bakanıyla görüşme yaptığını söyledi. Az buz bir şey değil bu; sıkı bir çalışma, yoğun bir gayret sarf ediyor Türkiye.

Dışişleri Bakanı geniş bir istişare toplantısı yapar da konu Avrupa Birliği sürecine gelip dayanmaz mı? Tabii ki aslî konulardan biri AB süreciydi. Bakan Bey'e eleştiriler de yöneltildi; yol gösterenler de oldu. Ali Bey gayet kibar, eleştiriye açık bir insan; notlar aldı, cevaplar verdi, bazen de söylenenleri tasdik etti. Ulusal Program, hem kâğıda basılı bir şekilde hem de dijital ortamda katılımcılara dağıtıldı. Ve Babacan, bundan sonraki yol haritasını izah etti. Tek kelimeyle yararlı bir istişareydi; hem katılımcılar açısından hem de davet edenler zaviyesinden...

AB sürecinde "heyecan eksikliği" sık sık dile getirildi ve hükümetin politikalarına ağır tenkitler yapıldı. Bunların bir kısmına katılmamak mümkün değil; keşke Türkiye iç siyasî çalkantılardan ve işe yaramaz polemiklerden yakasını bir an önce kurtarsa ve AB gibi geniş ufuklu projeler üzerine daha çok kafa yorabilse. Ancak, ısrarla yapılan bir yanlışa da değinmek

gerekiyor. Bazı kesimler AB sürecini hükümet(ler)in tek başına götürmesi gereken bir politikaymış gibi davranıyor. Bazı sivil toplum sözcüleri, AB konusunda hükümete adeta amele muamelesi yapıyor. Onları dinlerken sanıyorsunuz ki AB, sadece hükümetlerin işi. Tabii ki süreci iktidarlar yönetecek ama ona destek vermesi gerekenler yeterince doğru politikalar geliştirdi mi? Mesela AB sürecini engellemek isteyen statükoya sivil toplum yeterince karşı çıktı mı? "AB, bir hükümet değil devlet politikasıdır" demesine rağmen bu yolu sürekli tıkamaya çalışan muhalefet partilerine AB destekçisi sivil toplumdan tek satırlık itiraz yükselmiyor.

Dünkü gazetelerin birinci sayfalarında bu çelişkileri unutturacak ortak bir fotoğraf kullanıldı. Cumhurbaşkanı Abdullah Gül ile anamuhalefet lideri Deniz Baykal, samimi ve sevecen bir edayla Söğüt'teki programda görünüyorlardı. İşte bu! Farklı düşünmek ille de kavga etmek anlamına gelmiyor ki. N'olur farklı düşüncelerin zenginliği içinde birlik-dirlik mesajları verilebilse, bencillik ve parti popülizmi kokan tavırlar bir kenara bırakılabilse... Böyle bir atmosfer yakalandığında Babacan'ın sırtındaki yük de hafifler, fena mı olur?

### A.1.2 Sample Negative Turkish Political News Data

Yılın son günü devlet kendi vatandaşlarını öldürdü. Bir aile neredeyse tamamen ortadan kaldırıldı. Ardından elimizde kalan sadece kuru bir 'özür'... (o da tam dilenmedi ya...)

Böyle bir özrün faydası var mı?

Kırıp, döküp, öldürüp, yıkıp ardından 'kusura bakma' demeyi erdem mi sayacağız?

Açık konuşalım;

Bizden olmayanı öldürmek bu ülkede uzun zamandır bir yöntemdi ve bu alışkanlık hala sürmekte.

Bunun ardında bir siyasal irade var mı?

Var. Geleneği çok eskilere dayanan bir irade üstelik!

Kürt sorunu üstüne açık konuşmayı beceremedik. Eteğimizdeki taşları dökmeyi de!

Kim suçlu?

Herkes olabilir belki, ama orda can veren insanlar değil!

Kolay unutuyoruz.

Kolay unuttuğumuz kadar, kolay ölüyoruz.

Van'da türlü biçimlerde öldük.

Unuttuk.

Zindanlarda öldük.

Çıtımız çıkmadı.

Kocalar, sevgililer, babalar, ağabeyler dövdü, öldürdü.

Yüzümüzü çevirdik. Onurumuza, haysiyetimize kurşun sıkıldı, öldürüldük.

Üç maymunu oynadık.

'Bana dokunmayan yılan bin yıl yaşasın' dedik.

Yılan dokundu fark etmedik. Öldük.

Geride timsah gözyaşları aktı çoğu zaman.

Mertçe dövüşmekten vazgeçtik önce.

Sonra unutmayı öğrendik.

Şimdi bedavaya ölüyoruz.

Kötü alışkanlıklar edindik.

## A.2 Sample English Political News Data

### A.2.1 Sample Positive English Political News Data

Although Turkey, both at the level of the state establishment and the people in the street, is happy to say goodbye to Nicolas Sarkozy, the government may not immediately reverse its decision to partially suspend bilateral ties with France, which was made in response to Sarkozy's obsession with stigmatizing Turks using the Armenian issue.

This is not to say that the Turkish government is unwilling to normalize ties. On the contrary, Ankara will genuinely be looking to use this occasion to break the ice with Paris, which could also relieve the deadlock in Turkey's EU accession process. But Ankara may choose to wait and see to what degree France's new president, François Hollande, will insist on delivering on his promise to revive the law criminalizing the denial of Armenians' claims of genocide, despite the verdict of the French Constitutional Council.

It would be naive to expect Hollande to officially state that he will drop the matter. But a message to Ankara, that this issue will not be a priority, coupled with a clear intent to improve relations with Turkey, might suffice for the Turkish government to ease ties.

I wouldn't be surprised to see a very positive message of congratulation sent from Çankaya, which will be answered with equal warmth. This could be followed by a brief meeting between Turkish President Abdullah Gül and Hollande during the NATO summit in Chicago at the end of May.

So far the messages Hollande is sending leave room for optimism. Following the decision of the Council, Hollande promised to take up the issue of penalizing the denial of Armenian genocide claims, but emphasized that he would not be in a rush. Interestingly, in his statement he also addressed the Turks of France, saying they were wrong to think that the decision was directed against them.

This, according to Turkish and French experts, is the first time a high-level French politician has openly talked about the sensitivities of Turks living in France. For that, we can thank Sarkozy (as well as the Armenian diaspora in France), for it was his last-minute initiative to court Armenian votes that mobilized Turks living in France. Apparently many Turks with French citizenship rushed to their municipalities to register to vote. What's more, they have also been very active in the Socialist Party ranks, so much so that in the legislative elections set to take place in a month's time, more Socialist candidates, supported by a Turkish base, are expected to enter Parliament. They will be sensitive to their Turkish electorate, and this will in turn become an additional factor determining Hollande's stance on the Armenian issue.

Experts believe that it will be legally difficult for Hollande to challenge the decision of the constitutional council. But independent of that issue, the Armenian question will continue to be a headache in Turkish-French relations, because the Armenian Diaspora in France will commemorate the 100th anniversary of the 1915 tragedy under the the new Socialist government. However, despite the shadow of the Armenian issue there is room for optimism. First, the relationship will be cleared of the hostile rhetoric both sides have been using recently. And second, there is a high probability that Hollande will reverse Sarkozy's decision to suspend negotiations on five of Turkey's EU accession chapters. We should note that EU ambassador to Turkey Jean Maurice Ripert is a close friend of Hollande. That said, Turkey should be prepared for more vocal criticism from Paris on fundamental human rights issues.

### A.2.2 Sample Negative English Political News Data

The issue of journalists who lost their jobs because of their political stance and whose situation was brought up by deputies from the European Union (EU) constituted one of the most interesting topics in last week's meeting of the EU-Turkey Joint Parliamentary Commission in Istanbul.

Richard Howitt of Britain's Labour Party, also a member of the European Parliament's Socialist Group, voiced the most critical response, and this he did by specifically pronouncing the names of Nuray Mert and Banu Güven.

Speaking at this joint forum that brings together the representatives of the Turkish Parliament and the European Parliament, Howitt said the repression of press freedoms in Turkey had reached a worrying level and that opposition journalists were either placed under arrest or lost their jobs, while laying the blame squarely on the government's shoulders. Howitt expounded that Nuray Mert, a columnist for the daily Milliyet, had lost her space in the paper and that Banu Güven, a news show producer and host at the broadcasting station NTV, was forced to resign. "In fact, they were both journalists who had lent their support to the government in the past," he said.

"However, there are also a number of shortcomings that Turkey needs to tackle swiftly. The commission at various occasions expressed strong concern as regards respect for fundamental rights, and in particular freedom of expression and detention on remand," said Jean Maurice Ripert, the EU's new ambassador in Ankara, while expounding on the EU's position.

Bağış: Mert ought to be able to defend her ideas

EU Chief Negotiator and State Minister Egemen Bağış, who was present on the first day of

the meeting, responded to these criticisms.

“Ask that question to Nuray Mert and those who took her column away. How do I know? It must be either because her columns go unread or because her bosses dislike the ideas she entertains. This, however, cannot have anything to do with us or the government. If you draw a connection between [her losing her column] and her criticism of the government, then this argument is falsified by the continued occupation of columnists who are far more fiercely critical [of the government] than her,” Bağış said.

Bağış thus concluded his remarks: “Even if I disagree with her ideas, I attach great importance to Mert’s freedom to defend her ideas.”

Just as these talks were underway at the commission’s meeting, weekly British magazine the Economist also brought up Nuray Mert’s situation. Nuray Mert was “fired from her job” due to her dissident views, the Economist wrote.

A new problem in the West’s point of view

What do all these debates, writings and the close interest shown by Western embassies toward Nuray Mert indicate?

It seems the situation of dissident journalists is entering the agenda of European institutions, and the European public in general, as a new item in the list of the problems of Turkish democracy, after the case of arrested journalists. Apparently, we are going to hear the names of Ece Temelkuran, Mehmet Altan, Nuray Mert, Banu Güven and others rather frequently in the coming period.

The gaze then turns, of course, toward the government in wake of all these criticisms. It constitutes a separate matter for consideration the degree to which the government has any direct responsibility in these arrangements. Even in circumstances where the government does not bear any direct responsibility, the drawing of a connection between such practices that target journalists and the scope and the breadth of the freedom of expression in the country is inevitable.

To some extent, such practices of this kind are also viewed as a consequence, a reflection, or a by-product of the general political climate that holds sway in a country. Moreover, to a significant degree, it is none other than the government that is liable for this climate through its style, rhetoric, behavioral patterns and sometimes its passivity.

Prime Minister Recep Tayyip Erdoğan is in the habit of underlining the importance he ascribes to the freedom of expression at every turn. “We would not consent to others being subjected to what was done to us,” he had said at a reception of the daily Zaman last month. “We have been struggling to construe a milieu where everyone [can] speak freely in the language of their choice, and where no one feels the threat of repression hanging over them. We hold no doubts about our ideas, our beliefs and sense of right and wrong. We thus have no fear of anyone else’s ideas or thoughts, nor would we obstruct the freedom of expression. And we would not permit those who do want to obstruct it either,” he had added.

What does Erdoğan, who reminds us at every opportunity that he had “fallen” because of a poem he read, feel when it is journalists who “fall from their columns and programs”?