EMOTION DETECTION IN NOVELS

BİLGE BAŞDEMİR

JANUARY 2015

EMOTION DETECTION IN NOVELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES OF
ÇANKAYA UNIVERSITY

BY
BİLGE BAŞDEMİR

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

JANUARY 2015

Title of the Thesis: **Emotion Detection in Novels**

Submitted by **Bilge BAŞDEMİR**

Approval of the Graduate School of Natural and Applied Sciences, Çankaya University.

Prof. Dr. Taner ALTUNOK
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Müslim BOZYİĞİT
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assist. Prof. Dr. Gönenç ERCAN
Supervisor

Examination Date: 29.01.2015

Examining Committee Members

Assist. Prof. Dr. Gönenç ERCAN	(Çankaya Univ.)	
Assist. Prof. Dr. Gül TOKDEMİR	(Çankaya Univ.)	
Assoc. Prof. Dr. Ersin ELBAŞI	(İpek Univ.)	

STATEMENT OF NON-PLAGIARISM PAGE

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	:	Bilge, BAŞDEMİR
Signature	:	
Date	:	29.01.2015

ABSTRACT

EMOTION DETECTION IN NOVELS

BAŞDEMİR, Bilge

M.Sc., Department of Computer Engineering

Supervisor: Assist. Prof. Dr. Gönenç ERCAN

January 2015, 45 pages

Emotion is a feature unique to humans and is considered to be very difficult to imitate by a machine. A significant portion of printed sources is now being developed in electronic format and many people are following these resources in digital format with the development of technology. Determining the emotion of the text will enable new opportunities for the presented resources in electronic presentation whether voice or text. In this thesis, emotion detection is studied on two different author's novels. Turkish novels were not targeted in emotion detection in the literature before. Four machine learning methods are applied and compared on these novels which are determined as a dataset in this study. Same algorithms are applied to both novels separately and as a combination of both novels. The learning method is used for emotion detection in two different author's books in order to test if learning in emotion detection is dependent on the book and if effectiveness will drop due to this dependency.

Keywords: Emotion Detection, Machine Learning, Data Set

ÖZ

ROMANLARDA DUYGU TESPİTİ

BAŞDEMİR, Bilge
Yüksek Lisans, Bilgisayar Mühendisliği Anabilim Dalı
Tez Yöneticisi: Yrd. Doç. Dr. Gönenç ERCAN

Ocak 2015, 45 sayfa

İnsanlara özgü bir özellik olan duygunun makineler tarafından taklit edilmesinin çok zor olduğu kabul edilir. Teknolojinin gelişmesi ile birlikte basılı kaynakların önemli bir kısmı artık elektronik formatta geliştirilmektedir ve birçok kişi bu kaynakların takibini dijital ortamda yapmaktadır. Bu kaynakların ister ses, ister metin olarak elektronik ortamda sunumunda, metinlerin duygusunun hesaplanması yeni olanaklar sağlayacaktır. Bu tezde farklı yazarların romanları için duygu tespiti analizi üzerine çalışılmıştır. Türkçe romanlar duygu tespiti için daha önce kullanılmamıştır. Veri seti olarak belirlenen bu romanlarda dört farklı makine öğrenimi metodu uygulanmış ve sonuçlar karşılaştırılmıştır. Duygu tespiti için kullanılan öğrenme yöntemlerinin iki farklı kitapta hatta yazarda, o kitaba özel bir öğrenim yüzünden başarımın düşüp düşmediği test edilmiştir.

Anahtar Kelimeler: Duygu Tespiti, Makine Öğrenimi, Veri Seti

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Assist. Prof. Dr. Gönenç ERCAN for his supervision, special guidance, suggestions, and encouragement through the development of this thesis.

I am very grateful and would like to thank my mother Emire BAŞDEMİR, my sister Berna BAŞDEMİR and my father Zeki BAŞDEMİR for their invaluable patience.

I would like to thank all my judges especially Özkan YALÇIN for attending to my polls.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
NLP	Natural Language Processing
LDC	Linear Discriminant Classifiers
k-NN	K-nearest Neighbor
GMM	Gaussian Mixture Model
NB	Naive Bayes
VSM	Vector Space Model
SVM	Support Vector Machines
ANN	Artificial Neural Networks
HMMs	Hidden Markov Models
ML	Machine Learning
ISEAR	International Survey on Emotion Antecedents and Reactions
CNB	Complement Naive Bayes
SNoW	Sparse Network of Winnows
DT	Decisions Tree
RF	Random Forest
IB1	1 Nearest Neighbor
ANEW	Affective Norms for English Words
ME	Maximum Entropy
ARFF	Attribute Relationship File Format
DM	Data Mining
UL	Unsupervised Learning
SL	Supervised Learning
TF	Term Frequency
IDF	Inverse Term Frequency

IMDB	Internet Movie Database
WEKA	Waikato Environment for Knowledge Analysis

CHAPTER 1

INTRODUCTION

Emotion is an important trait specific to people. People can take important decisions with their emotions. Emotions vary depending on their daily lives and psychology. This ability changes according to a person's feelings.

Emotion is shaped by human's creativity, social interaction, learning, teaching and any occurrences in their life. It is commonly believed that emotions are one of the most challenging abilities for machines to mimic.

Technology is improving day-by-day, hence, people depend more and more on technology. Humans do everything with technology to facilitate their work and save their time from their daily routines. Internet is indispensable in our life along with the development of the technology. It was difficult to collect all information or comments following all media channels and publications.

Due to this development more than one internet media resources emerged, such as blogs, twitter, facebook, dictionary pages, forum pages and company social web pages. New media resources can be initiated in a day, even in a minute.

These media resources have lots of users and their comments about any title in the web sites. To follow, save time and take control of technology, researchers found Emotion detection analysis.

For example we may want to predict the effect of news on people using comments of the news retrieved from the newspaper's web page. Emotion detection can be used to predict people's feelings about the news. Due to this reason, Emotion detection is an increasingly popular subject in Artificial Intelligence (AI) research.

Emotion detection is an active research topic, attracting interest from Artificial Intelligence (AI) researchers. These studies are done for text, speech, facial expressions, according to word's emphasis, and people's psychology [1, 2]. Emotion detection on text are studied by many researchers especially focusing on microblogs such as twitter. Good results have been achieved in emotion detection from texts [2].

While most of the research on emotion detection target microblog data or user comments, only few previous efforts target novels as we done in this study. We will find emotions of each sentence in the text. At the end of this study, it will be possible to develop text-to-speech and e-book readers that can act depending on the detected emotion. For example a text-to-speech engine can set the tone according to the detected emotion, or an e-book reader can change the background according to the emotion in the novel.

Studies of emotion detection use Natural Language Processing (NLP), machine learning, computational linguistics and symbolic techniques approaches. Most frequently used classifiers are linear discriminant classifiers (LDC), k-nearest neighbor (k-NN), Gaussian mixture model (GMM), Naive Bayes, Vector Space Model, support vector machines (SVM), artificial neural networks (ANN), decision tree algorithms and hidden Markov models (HMMs). Used emotions are Joy, Fear, Sadness, Anger, Disgust, Trust, Surprise, Anticipate, Shame, Guilt and Neutral for available studies (Tomkins, 1962; Izard, 1977; Plutchik, 1980; Ortony et. al., 1988; and Ekman, 1992).

Research on emotion detection mostly uses English text datasets. Literature lacks research which uses Turkish data set. In this thesis, we fill this gap by creating and using a Turkish text data set. Data sets are created from two Turkish novels (Agatha CHRISTIE "Şampanyadaki Zehir" and Zülfü LİVANELİ "Serenad").

In 1992 Paul Ekman (psychologist) worked on classification of emotions. He created a list of six basic emotion types. These basic emotions are; anger, disgust, fear, joy, sadness and surprise [3]. In this thesis classifications of six emotions, which are joy, fear, worry, sadness, anger and neutral are handled in both these two novels. Five judges answered our data set. Every judge read sentences and decided which emotion is suitable for the sentence according to their feelings.

Classification algorithms are applied by using supervised & unsupervised machine learning techniques in this study. NB, J48, SVM and KNN classification algorithms were used. All results of algorithms and novels were compared.

1.1 Outline of the Thesis

This thesis is structured with five chapters. We gave all details of emotion detection in machine learning and results of different algorithms.

Chapter 2 starts with a literature survey of emotion detection methods, data sets and their differences from our study. Meaning of basic terms and schema about our title. Basic emotions are given which were studied in emotion detection. Also information about Machine Learning (ML), commonly used algorithm types and ML classification methods are given.

Chapter 3 contains information about our study's data set and its example also feature types of this thesis.

Also details of our data are given in this chapter. Chapter 4 presents the experimental results and comparison between classification methods.

Finally we conclude with a discussion of our work and possible future work.

CHAPTER 2

LITERATURE SURVEY

Aim of this study is: detecting the emotion of sentences in Turkish novels. The terms emotion and sentiment are defined in Section 2.1. In Section 2.2 introduces some basic notions about Machine Learning (ML) algorithms. Section 2.3 describes the classification algorithms evaluated in our experiments. General information and descriptions about this emotion analysis is presented in Section 2.4, followed by an analysis of Turkish & English emotion detection systems on text documents.

2.1 What is the Emotion & Sentiment?

2.1.1 Emotion

Emotion is the fundamental characteristics of human being that makes them different from machines. Word's meaning & emphasis, punctuation is explaining state of emotions.

According to Kleinginna [4] two agreed-upon appearance of emotion available which are:

1. Emotion is a reaction to assume events. All events include people' concerns, goals, thoughts, needs and so on.

2. Emotion covers physiological, affective, behavioral, and cognitive of all living beings.

There are two kinds of models for emotion separation: the categorical model and the dimensional model. These two models are used to represent all different emotions. In the categorical model primary emotions are defined to represent all emotions. These six primary emotion categories are anger, disgust, fear, joy, sadness and surprise [5]. In the dimensional model each emotion is represented as a dimension and all emotions depend on each other with common emotions in the high dimensional space [6-8].

In Figure 1 the dimensional model is shown, which is called as the emotion wheel [9].

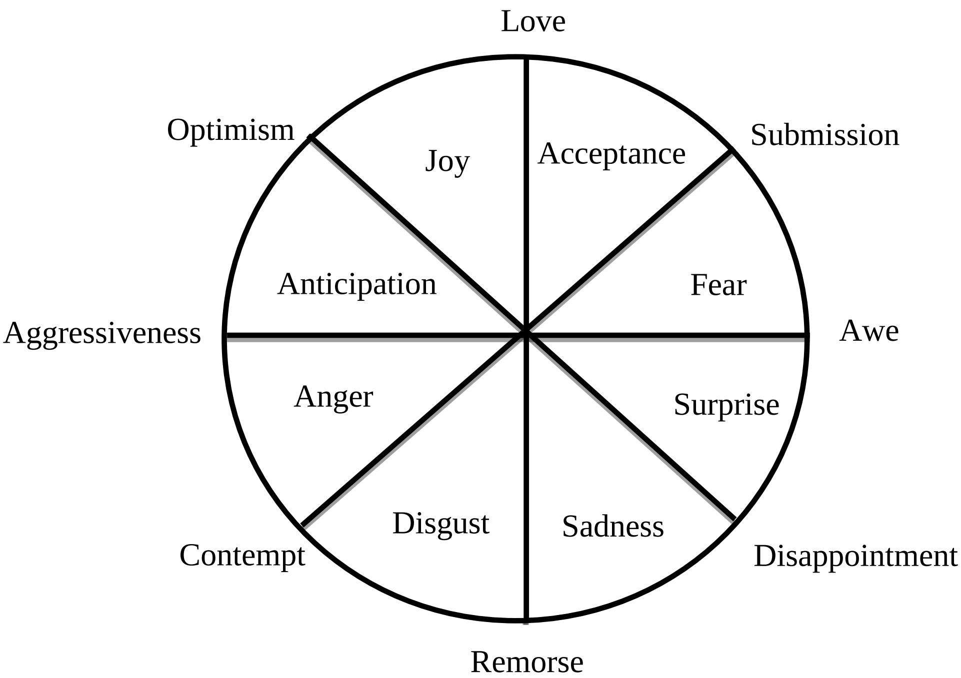

Figure 1 Plutchik's emotion wheel

Paul Ekman defined the most basic universal emotions as anger, disgust, fear, joy, sadness and surprise. Six emotions were selected by Paul Ekman because these emotions can be defined for all people. This list was extended in face recognition by Ekman for his other studies to understand more kinds of emotions. All emotion analysis's studies worked on these basic emotions because of that we also studied on basic emotions.

2.1.2 Sentiment

Sentiment is a more general concept compared to emotions. It includes attitudes, emotions, and opinions. Using NLP and machine learning methods sentiment of texts can be identified. Pang, B. and Lee, L. introduced a seminal work in this field [10]. Sentiment analysis methods are similar to emotion detection's methods.

2.2 Machine Learning

Computers are in every moment of people's lives because of that, in the hospitals, schools, government institutions, universities and many more places are keeping all transactions (events, data, and records of camera) in a database. These databases are specific to these places and are typically large in volumes. There is much important hidden information in these data, using the past data we can estimate future data [11].

Important and meaningful information are uncovered from data by the use of data mining algorithms. Figure 2 shows basic steps of Data Mining [12].

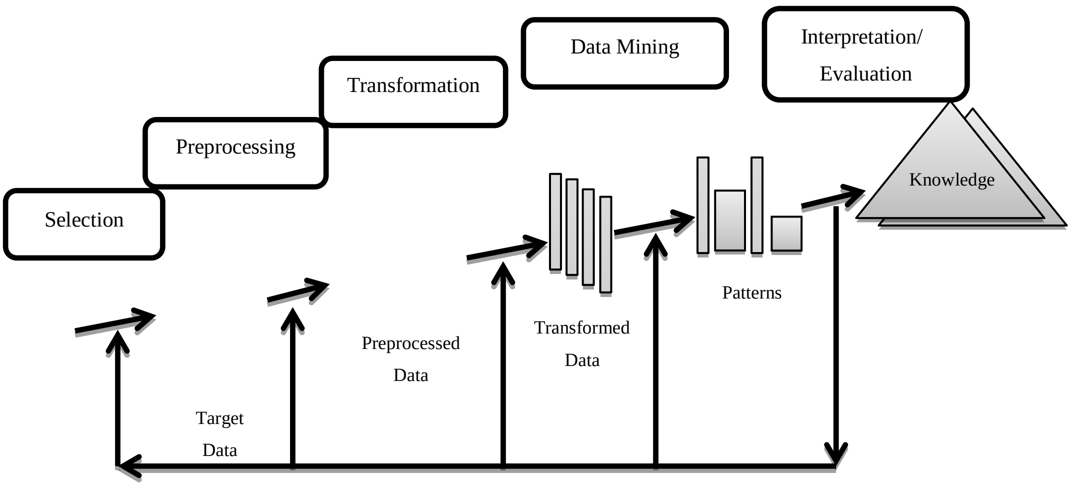

Figure 2 Data mining process

Analyzing and processing these large amounts of data manually is impossible, it takes lots of time. To solve this problem Machine Learning (ML) methods are utilized. To realize data mining we need analyzing programs (for example WEKA and Rapid Miner).

ML is doing modeling and the model is based on data from the environment of a given problem. In ML many approaches are proposed, providing a variety of algorithms.

There are many suggested machine learning methods available where some of them are more suitable a given problem than others. According to No Free Lunch Theorems; different machine learning algorithms may have different success in different problems [13].

ML algorithms can be divided into several categories according to the intended result. The algorithm types commonly used are Supervised Learning (SL), Unsupervised Learning (UL).

2.2.1 Unsupervised learning (UL)

An UL method uses a model input which uses unlabeled data. In UL which object in which class are not known and furthermore number of classes may not be known. UL is the process of finding hidden structure in raw data. Clustering methods are examples of UL. We applied some methods of UL in machine learning to do classification to our data in WEKA.

2.2.2 Supervised learning (SL)

Supervised machine learning is the search for algorithms that reason from externally supplied instances to produce general hypotheses, which then make predictions about future instances. The goal of supervised learning is to build a concise model of the distribution of class labels in terms of predictor features [14].

In SL the number of classes and which objects are in which classes are known. It uses the data and the results. These data and results are known in advance. SL aims to create a function using these results. Previously known training data sets output is uploaded to the system with SL.

In this study, SL machine learning methods are used, as we learn a classifier from labeled. We know which instance is in which class in SL using these information after that, system is doing learning. In this study, training was conducted based on the data. SL needs labeled data set and details of the data set are explained in 3.2. The steps of our study are the same as in Figure 3 [14].

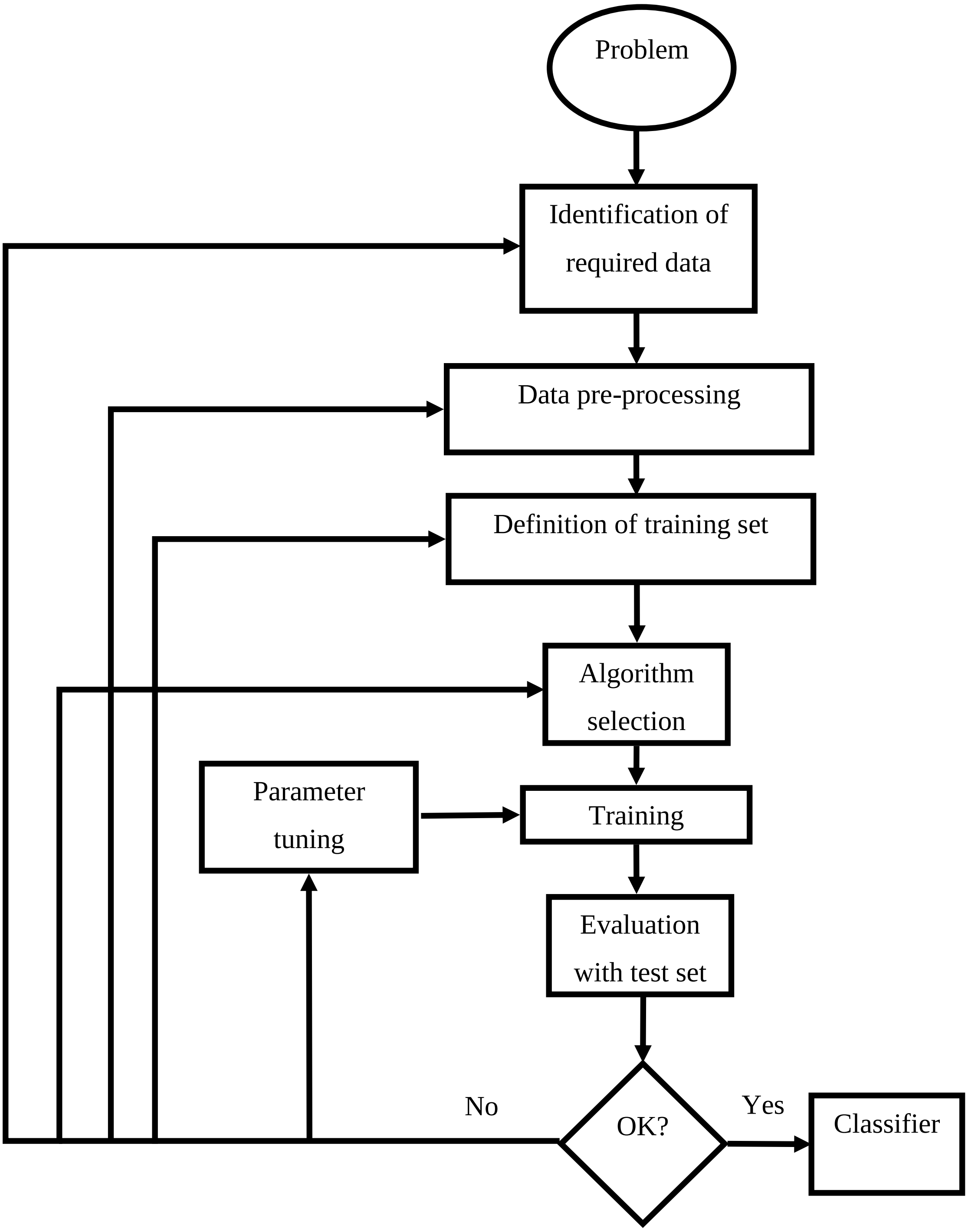

Figure 3 SL's processes [14]

We divided the system into two parts: training set and test set to measure the success of the system. In our study, system is learning from the training set. The success of the trained system is calculated on the test set.

Formally emotion detection can be defined by two lists, where the first is the list of sentences S= {S_1, S_2 … S_n} and the second is the emotion of corresponding sentences (n=2145 for each data set) E = {e_1, e_2…e_n }, where $e_i \in$ C= {c_1, c_2…c_k} and C is the set of emotion categories.

While in our research the emotion categories are 6, as defined in Ekman [3] it is possible to extend this representation to further support more emotion categories. In this study we are trying to locate the E list for a given document of S list. Using the training set a machine learning which maps the sentences to emotion categories is built. After that we try to estimate the E list for the test sentences.

2.3 Machine Learning Classification Methods

This section introduces the classification algorithms used in our experiments. We perform cross validation to test feature type and classification methods. In K-Fold cross validation the whole dataset is divided into equal-sized groups of N. One group is uses for testing and the other groups are used for training. This process is repeated N times. Average of these N experiments is reported as the overall success.

We applied Naive Bayes (NB), J48, KNN and SVM methods in our study because many researchers applied these methods to analyze emotion detection in the related work. These four methods are more popular for the classification of emotions.

2.3.1 Naive bayes (NB)

NB is a supervised classification method in machine learning. NB assumes that the attributes are independent from each other. It is a kind of probabilistic classifier which based on applying Bayes' theorem. NB classifier is a simplified version of the independence of the Bayes theorem proposition. Generally, NB is used for the text categorization in machine learning and receives satisfactory results [15].

Bayes theorem provides inferences using the collected data about a system, which system is to be obtained. Also it helps to estimate current situation when a new data is added to available system. Bayes’ theorem equation is:

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} \tag{2.1}$$

C = Class

D = Document

P (C|D) = probability of D to be in the C

P (D|C) = probability of C to be in the D

P (C) and P (D) = known probability of C and D

With this calculation we can estimate which document is in which class.

In NB classification all events are important and independent (one event not including any information about other event) [16]. Classification is made with the help of statistical methods in NB.

NB’ theorem equation is:

$$P(A|C1, C2, \ldots . Cn) = \prod_{k=1}^{n} \frac{P(C\mathrm{i}|A)P(A)}{P(C\mathrm{i})} \tag{2.2}$$

C = Class

A = Document

$P(A|C1, C2, \ldots . Cn)$= probability of document A likely to be of any class

P (A) = probability of A

$P(C\mathrm{i}|A)$= the probability of C is given A

$\prod_{k=1}^{n} P(C\mathrm{i}) =$ multiplied by the probability of belonging to any class

2.3.2 Support vector machine (SVM)

SVM is one of the supervised learning algorithms it usually is used in text classification. It is a simple and an effective classification algorithm. Labeled training data is uses SVM and creates a vector model to separate groups for classification. These groups are separated by a hyperplane, so need to find the hyperplane.

SVM determines how to draw this hyperplane and to learn data which is close to hyperplane. The location of this hyperplane will be drawn taking place for from the members of two groups.

SVM can classify data sets as linear as well as non-linear [Figure 4-Figure 5].

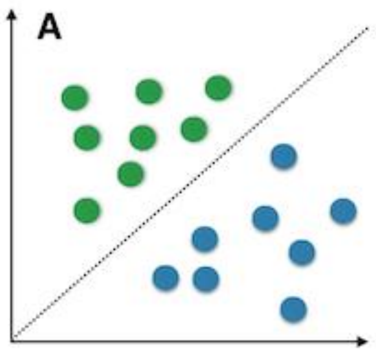

Figure 4 Linear data

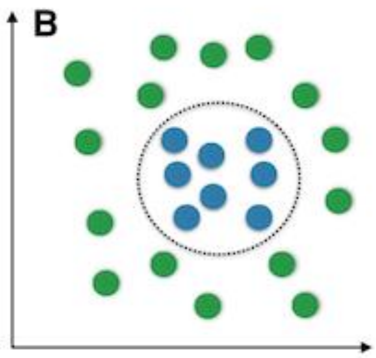

Figure 5 Non-linear data

SVM, linear plane defined by the support vectors and support vector tries to find the maximum range. Figure 6 presents the data of two classes in the vector space. Assume that blue circle's class is Joy and red circle's class is Fear. Each circles has specific attributes combining x and y values that means each data is occurring by two values in a vector space. In this scheme it is enough to separate data by straight-line (so this scheme's hyperplane is a straight-line), but if three values for each data this time there should be a plane to separate data [17].

If four or more than four values for each data this time there should be a hyperplane to separate data.
$w.x - b = 1$ and $w.x - b = -1$ are are hyperplanes which are called support vectors and dividing optimal hyperplane at the edges of the margin. SVM's objective

is to find the edge to maximize the distance between the support vectors. The edge of a hyperplane which is located equidistant from the support vectors, classification is the best way separating hyperplane of each other.

$$\sum_{i=1}^{l} f \propto i \;\, yi \; xi \tag{2.3}$$

Where “*l* “=number of the support vector number,

“xi “is the support vector,

“yi ϵ 1,-1” is the class sign, and

Constructing the optimal hyperplane is equivalent finding nonzero " $\propto i$"

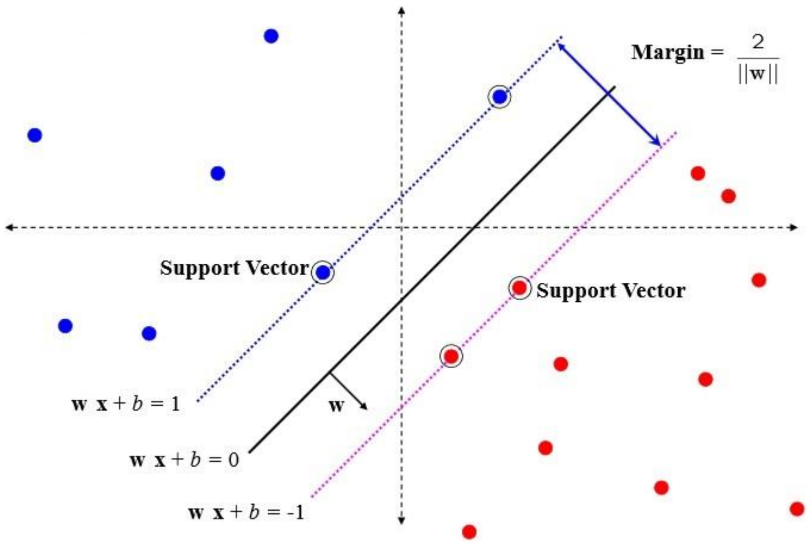

Figure 6 Linear separating using hyperplane

2.3.3 J48 (Decision Tree (DT))

J48 is a decision tree that is a predictive machine learning model. It creates a new instance model built from the attributes of available data.

Probability of a result is multiplied of all factors which are affecting to provide results’ probability in J48. It is based on C4.5’s algorithm which is improved by Ross Quinlan in 1993 [18].

Entropy is used to evaluate the classification ability of a feature in ML. Entropy comes from information theory. Higher entropy means it is including more corrupted data. Entropy values are used to draw the decision tree. DT divides the data set by a feature which has the highest information gain value. In a way, we decide the attribute which provides the most benefit [19].

$$\text{Entropy} = \sum_i -p_i \ . \log_2 p_i \quad (2.4)$$

p_i = Probability of class

An example of J48 is represented in Figure 7 which has three nodes. Each node (outlook, humidity and windy) represents a feature of the available data. Applying a divide and conquer logic, data is partitioned with respect to attribute's values until each node has only elements from one class. Each branch (sunny, overcast, rainy, high, normal, false, and true) represents a value for feature. DT algorithms analyze set of cases and the instance clusters within the training set.

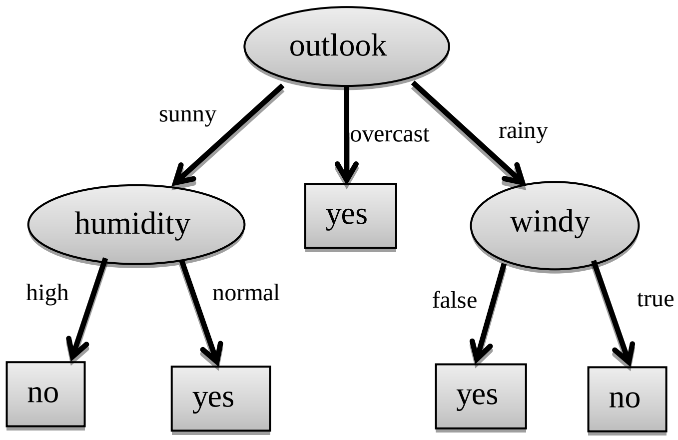

Figure 7 Example of decision tree

2.3.4 K-Nearest neighbors (KNN) algorithm

KNN algorithm is a classification algorithm based on supervised and sample (instance). Sample to be tested on training set is compare with each training instance. In order to determine a class for a test instance, the K nearest sentence from the training set are selected. Aim of KNN is to classify a new vector [20].

Class with the most exemplary in selected samples is assigned to the test instance. The important thing is to determine in advance how to calculate the similarity and the features used in the process.

$$\text{dist}\,(X_1, X_2) = \sqrt{\sum_{i-1}^{n} (X_{1i}, X_{2i})^2} \tag{2.5}$$

Assuming that X vector consisting of n attributes ($X_1, X_{2,\dots\dots\dots} X_n$). Calculation is done using the above equation. This equation is called Euclidean distance, which calculates the distance between two samples.

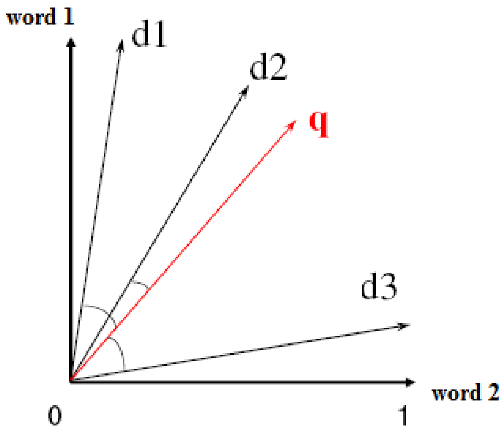

Figure 8 Two-dimensional vector spaces

q is the vector that we want to find for the class.

d_1, d_2 and d_3 represent consist of training instance vectors.

In Figure 8, each dimension corresponds to a word, so each text can be expressed in the vector space. Using K-NN algorithm these texts can be determined how much they resemble each other in the vector space. Namely, texts that are nearest to each other can be identified.

In Figure 9 two classes are represented in a vector space. Each instance has a distance to class which it belongs to. Green instance's class is not known. Selecting nearest neighbor by calculating the Euclidean distance we can find class of this instance.

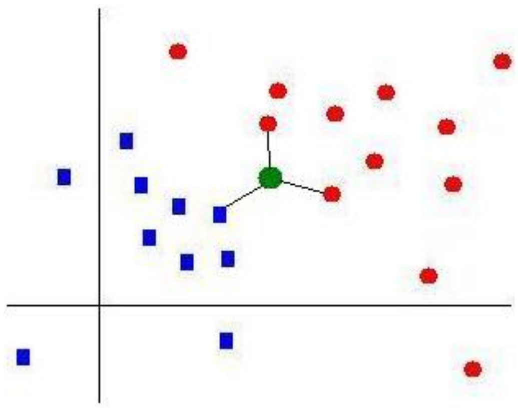

Figure 9 KNN scheme in a vector space

Also, similarity calculation can be used in KNN to find class of instance.

$$\text{sim}\ (d_i, q) = \frac{d_i\ .\ q}{|d_i\ |\ .\ |q|} = \frac{\sum_j W_{i,j}\ x\ W_{q,j}}{\sqrt{\sum_j W_{i,j}^2}\ .\sqrt{\sum_j W_{q,j}^2}} \tag{2.6}$$

$W_{i,j}$ = Weight of terms in document

d_i = Training document vector

q = is the vector that we want to find the class

2.4 Emotion Analysis in Related Works

Emotion analysis is a popular subject, because many researchers propose different methods. In particular, emotion classification of text has been studied [1]. In the recent scientific studies related to emotion analysis, there are two basic approaches. Namely, machines learning based [21] and dictionary-based [22] emotion analysis.

Supervised machine learning analysis uses labeled text to learn to classify emotion. Machine learning algorithms are used to train the system with a portion of the data set. Dictionary-based analysis in a sense is a word weighting method classifies text using an emotion enriched dictionary [23].

Many researchers worked on the problem of Emotion Detection. Different emotion categories identified by different researchers are shown in Table 1. These emotions categories universally acceptable. In our study we used Ekman's emotions which are accepted by many researchers.

Table 1 Using emotions by Different Researchers

Tomkins (1962)	**Izard (1977)**	**Plutchik (1980)**	**Ortony et.al. (1988)**	**Ekman (1992)**
Joy Anguish Fear Anger Disgust Surprise Interest Shame	Enjoyment Sadness Fear Anger Disgust Surprise Interest Shame Shyness Guilt	Joy Sorrow Fear Anger Disgust Surprise Acceptance Anticipation	Joy Sadness Fear Anger Disgust Surprise	Happiness Sadness Fear Anger Disgust Surprise

Researchers used many different data sets for emotion detection implementations. Used data sets are: Twitter's tweets [24-26, 28], Turkish & English movies [27-30] and novels [31, 29]. We used Turkish novels as a data set in our study so we compared differences between our study and other studies targeting novels.

2.4.1 Emotion analysis for novels

Z. Boynukalın used Machine Learning (ML) algorithms for emotion detection. Four emotions are used in this study for the classification; these emotions are joy, sadness, fear, anger and neutral. Turkish children's tale and ISEAR (International Survey on Emotion Antecedents and Reactions) [32] are used as a data set. ISEAR is a project which is managed by Klaus R.Scherer and Herald Walbott. A total of 3000 people from 37 different countries participated in this project for emotion classification survey. ISEAR data set has been translated into Turkish language. With this study, machine learning methods are applied on Turkish texts and shown to give satisfactory results. Naive Bayes (NB), Complement Naive Bayes (CNB) & Support Vector Machines (SVM) classification methods are compared [31].

Boynukalın's study is more similar to our study. Five emotions are used and three judges labeled data set's sentences according to five emotions in Boynukalın's study. In our study six emotions are used and five judges labeled our data sets. They studied on ISEAR dataset to Turkish translations and Turkish Fairy Tales Dataset, but we studied on two novels as a dataset. Both studies' data sets are in Turkish, but different genres are used. In our study, we investigated classification of emotions' performance on two different authors' book.

Alm performed the first study in emotion detection in novels and fairytales [29]. Supervised machine learning is used with SNoW (Sparse Network of Winnows) learning architecture. In this study 22 fairy tales as preliminary data used for pilot study. They used a feature set of 30 features for their dataset, including the first sentence of the story, sentence length in words, verb count in the sentence, WordNet Affect [13] words, with a program written in Python programming language. In their study 2 types of classification is used. The first one is, classifying as positive, negative or neutral, which is sentiment analysis. The second one is classifying as emotional or non-emotional. They used seven emotions: angry, disgusted, fearful, happy, sad, and positively surprised and negatively surprised. A difference from our study is: they used seven emotions to label data by judges, they used 30 features and Alm's data set' language is English ours is Turkish language.

2.4.2 Emotion analysis for Twitter

In 2009 Go, Bhayani and Huang suggested an approach for classifying twitter posts into two classes as positive and negative [24]. This study can be used by individuals and companies that may want to research sentiment on any topic according to author's opinion. For this study, 800.000 positive tweets and 800.000 negative tweets were used. They analyzed 1.600.000 English tweets as a dataset. For this study unigram, bigram and combination of both were used to analyze the tweets. The result of this study is presented in Table 2.

Table 2 Used Algorithm Results

Features	Keyword	NB	MaxEnt	SVM
Unigram	65.2 %	81.3 %	80.5 %	82.2 %
Bigram	N/A	81.6 %	79.1 %	78.8 %
Unigram + Bigram	N/A	82.7 %	83.0 %	81.6 %
Unigram + POS	N/A	79.9 %	79.9 %	81.9 %

Çetin & Amasyalı [25] compared two data sets using Machine Learning processes. Comparing traditional methods and supervised method we used to learn which method is more effective in sentiment detection. They used Naive Bayes (NB), Support Vector Machine (SMO), Decision Tree (J48), Random Forests (RF), 1 nearest neighbor (IB1) algorithms. They used tweets belonging to two private companies in the telecom sector as a dataset, containing 6000 samples. These data sets are divided into 3 classes with positive, negative and neutral sentiments using manual procedure. They selected telecom sector because emotion analysis is important subject for companies with potential benefits in terms of commercial profit. The best performance is achieved when using SMO algorithm.

E. Akbaş [26] separated texts into subjects and according to his hypothesis subject information is important in emotion analysis. He used Turkish tweets as data set. Data sets were about tweets of 3 GSM companies operating in Turkey. Through Twitter API, Akbaş filtered and divided tweets according to subjects such as billing and Internet. Emotion detection is done using machine learning algorithms separately for each subject. They used a dictionary-based feature selection method in the learning process. In this study, texts are classified by making the weighting towards the negative than positive emotions.

In Şimşek’s study emotion detection is done using a dictionary-based approach. "Emotions in Social Psychology" is used to create the data set [33]. Happiness and sadness that specifies a set of 111 words was prepared as a query. The happiness value is given to these words between 1 and 9 considering Mechanical Turk data [34]. Approximately 1.9 million Turkish tweet data has been prepared using the Twitter API. Tweets texts have been found which are including 111 words. Vector model is applied using average value of the words. Happiness and sadness analysis was performed on texts [35].

Nielsen [28] introduced a dictionary-based Twitter mood in 2011. In this study Affective Norms for English Words (ANEW) data set was used. ANEW consists of 1034 words and which is an emotional weights feature words. Valence, arousal and dominance scores between 1 and 9 are given to 1034 words. ANEW’s emotional weights were obtained from a study of college students at the University of Florida. The average of the student values obtained from participants and emotional word weights are determined. Twitter mood of the text was calculated using the valence coefficients of ANEW words. According to the mood of America Twitter data shown on the map of America.

2.4.3 Emotion analysis for movies

The first basic study in emotion detection is done by Pang, Lee and Vaithynatham in 2002 [27]. User comments in Internet Movie Database (IMDB) are extracted to form the dataset for English Language. In this study researchers used Machine Learning (ML) methods. Using these methods, researchers classified reviews as positive, negative or neutral. However, classification algorithms used as SVM, Naive Bayes, Maximum Entropy. The best results are obtained when using unigrams. Most effective classification algorithm result was achieved when using SVM (%82.9).

Sevindi [30] uses Turkish movie comments as a data set for sentiment analysis. In this study three emotions are used which are positive, negative and uncertain. He defined emotion poles using different methods. In this study Machine learning and dictionary-based approaches are used and results of these approaches are used and compared. For machine learning methods C4.5, KNN, NB and SVM were used and N-grams's attributes are used: unigram, bigram, trigram as feature types. The best result in Machine Learning approaches for F-measure was taken from SVM algorithm which was 82.58 %. Result in dictionary-based was 59.69 %. Machine learning methods were producing more successful results for Turkish movie's comments. A difference from our study is number of using emotions and B. İbrahim compared machine learning and dictionary-based approaches used and compared, but we compared results of novels. Sevindi is used four algorithms as our study. In table 3 is giving summarizes of related works about basic information of emotion detection.

Table 3 Summarizes of Related Works

	Novels	**Twitter**	**Movies**
Data Sets	ISEAR Fairy Tales	Twitter Comments ANEW	IMDB Turkish Movie
Methods	NB CNB SVM	NB ME SVM J48 RF IB1	NB SVM ME C4.5 KNN

CHAPTER 3

RESEARCH METHOD

This thesis aim is to measure the success of emotion classification for two authors of the book. Section 3.1 gives information about the details of the data set built and used in this thesis. Section 3.2 gives details of the Feature Types. Basic information about WEKA is expressed in Section 3.3.

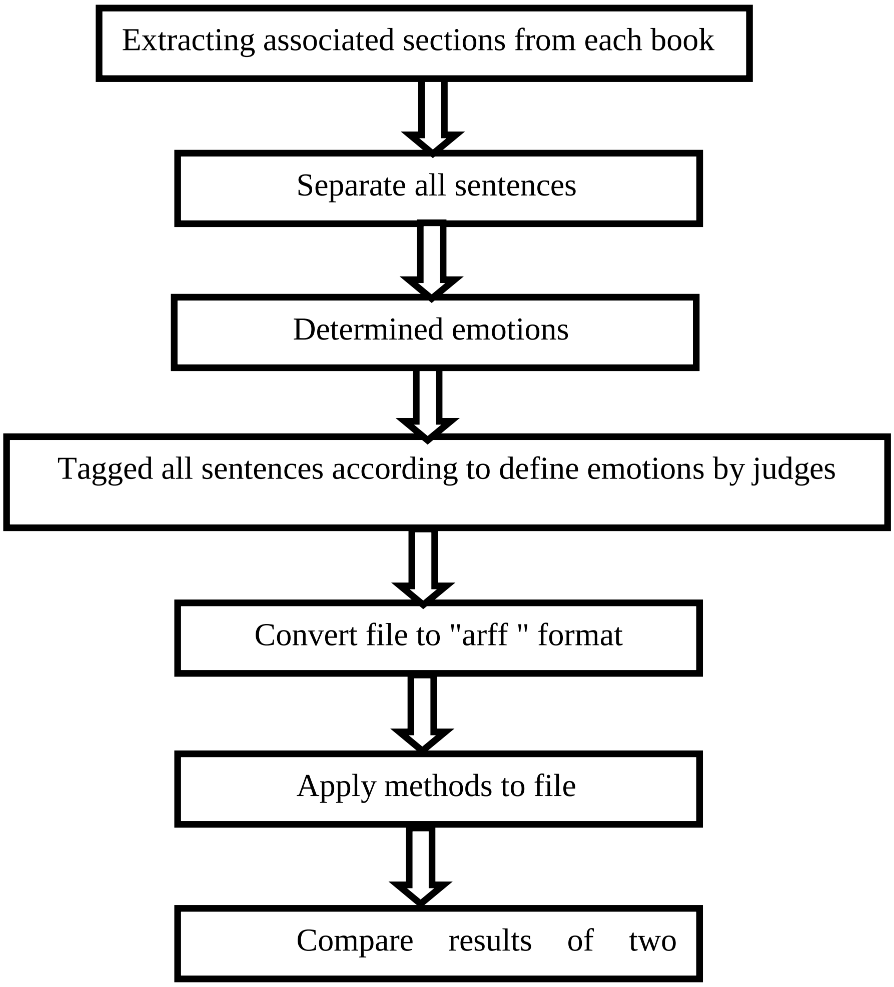

Figure 10 Overview of our system

As shown in the Figure 10 we did our study according to several steps. We have seven steps which summarize the work done on this thesis. In the first step we have determined two novels and its sections as a dataset, after that sentences are partitioned using these sections.

According to our researches we discussed basic emotions which are defined for emotion detection, and we decided which emotions we will use. In the fourth step judges tagged sentence by sentence according to emotions which are defined in the previous step. We convert our dataset to “arff” format after voting polls, and machine learning methods have been applied to find emotions of books. Finally we compared results of two novels to analyze which methods achieve better results than other methods.

3.1 Data Set

In this thesis two Turkish Novels are used to form two data sets. These novels are “Şampanyadaki Zehir” by Agatha CHRISTIE (first DataSet) and “Serenad” by Zülfü LİVANELİ (second DataSet). Subjects and authors of these novels are deliberately selected as different. Subject of first DataSet is fear and subject of second DataSet is love. Reasons of two different novels are selected is to compare and analyze emotion in different books. Our Emotion detection experiments are conducted using these two datasets.

Both datasets included 2145 sentences, yielding a total of 4290 sentences. Sentences are interdependent, i.e. taken from the sequential parts of the book. These sentences are numbered and are annotated by the human judges. The same procedure is used for both datasets.

In this study, six basic emotions are used as the target categories set C, namely they are joy, fear, worry, sadness, anger and neutral. Seed words are given to human judges in order to clarify the categories. This list is shown in Table 4. Judges considered this list; otherwise they took into consideration their feelings when they are reading the sentences.

For the two novels totally 5 people participated to label all sentences. Five people's results were combined and common emotions were selected for each sentence to analyze emotions of the books. For all sentences of data sets' emotions assigned according to majority vote from all votes.

Table 4 Seed Words List for Emotion Categories

Emotions	**Seed words**
Joy	Mutluluk, sevinç, neşeli, mutlu
Fear	Titreme, ürperme, irkilme, dehşet, korkutmak
Worry	Şaşkınlık, mucize, şaşırmak, heyecanlanmak
Sadness	Keder, mutsuzluk, hüzün, sıkıntı, üzüntü, umutsuzluk
Anger	Öfkeli, kızgın, sinirli, nefret dolu, sert, asabi
Neutral	Duygu yok

Two novel excerpts of 35 pages are prepared for this thesis. Results of the judges were digitalized to find the common emotion from all judgments for each sentence. Common emotion is shown in Table 5.

Second poll was prepared and sent to judges in the digital environment. Results were entered and common emotion decisions are determined for all sentences in a similar fashion to first poll.

Table 5 Result of Judges

Sentences	**Judge 1 Berna Başdemir**	**Judge 2 Bilge Başdemir**	**Judge 3 Özkan Yalçın**	**Judge 4 Kamil Çakmak**	**Judge 5 Fulya Göksu çukur**	**Common emotion**
Ama aslında gerçekten beklenmedik miydi?	fear	worry	worry	neutral	worry	worry
Daha önceden belirtileri görülmemiş miydi?	neutral	worry	worry	neutral	worry	worry
George'un dalgınlığı, garip davranışları...	neutral	neutral	worry	sadness	neutral	neutral
Başka türlü tanımlanamazdı bu.	neutral	neutral	neutral	neutral	neutral	neutral
Rosemary nasıl bir insandı?	neutral	worry	worry	neutral	worry	worry

Table 6 shows example of sentences and common emotions by judgesters of first DataSet.

Table 6 Examples of First Dataset Sentences with Emotions

Sentences	Emotions
Ama aslındagerçekten beklenmedik miydi?	worry
Daha önceden belirtileri görülmemiş miydi?	worry
George'un dalgınlığı, garip davranışları...	neutral
Başka türlü tanımlanamazdı bu.	neutral
Rosemary nasıl bir insandı?	worry
Ablam aşığıyla kaçmanın neye mal olacağına da aldırmıyormuş.	anger
Ve Rosemary de adam kadar kararlıymış...	neutral
İris ürperdi.	fear
Benim bu durumdan hiç haberim yoktu.	neutral
Hiç şüphelenmedim bile.	neutral
Rosemary'nin mutlu ve rahat olduğuna, George'la birbirlerini sevdiklerine inanıyordum.	neutral
Ne körmüşüm!	anger
Ablamın halini fark etmedim bile...	anger
Ama kimdi bu adam?	worry
Geçmişi düşünmeye, hatırlamaya çalıştı.	neutral
Rosemary'nin etrafında bir suru hayranı vardı.	neutral
Özel biri yoktu.	neutral
Ama olması gerekir...	neutral
Herhalde bu hayranlar sadece kamuflaj içindi.	neutral

Table 7 shows example of sentences and common emotions by judgesters of second DataSet.

Table 7 Examples of Second Dataset Sentences with Emotions

Sentences	Emotions
Zeki değildi ama benzerlerinin çoğu gibi kurnazdı.	neutral
Galiba zekâ ile kurnazlık ters orantılı.	worry
Biri azalırsa oburu artıyor.	neutral

Bu düşüncelere dalmışken, peşimize yapışmış bir otomobil olduğunu fark ettim.	worry
Ya bizim eskortumuz sanıyorlardı ya da farklı bir şey vardı.	worry
Tıkalı yollarda saatlerdir bekleyen binlerce araç sahibi bizi öfkeli gözlerle süzüyordu.	anger
Boston'da da trafik böyle mi profesör?	worry
Dalgınlığından sıyrılıp yumuşak bir sesle.	neutral
Hayır dedi.	neutral
İyi ki de değil, çünkü orada üniversitelerin böyle ayrıcalıkları yok.	neutral
New York böyledir herhalde!	worry
Evet, orası bazen yoğun ama yine de böyle olduğunu sanmıyorum.	neutral
Bu kadar araba nereden cıktı anlamadım.	worry
Benim zamanımda yollarda tek tük otomobil görülürdü.	neutral
Herkes işe tramvay ya da vapurla giderdi.	neutral
Köprüler de yoktu tabii.	neutral
Galata Köprüsü mü? Vardı.	worry
Hayır, Boğaz köprülerini kastediyorum.	neutral
Avrupa'yı Asya kıtasına bağlayan iki köprü.	neutral

Example of first Dataset's sentence is "Ablamın halini fark etmedim bile...". Agreement of judges for this sentence is anger. This sentence's emotion can be classified as a different emotion when it is considered alone, but our data set is a sequential part of a novel, so judges consider the context acquired from previous sentences (Ne körmüşüm!).

We measure of the agreement of judges' kappa statistics. Kappa is measure of how much judges agree or disagree [36].

$$K = \frac{RJ - P\mathrm{I}}{1 - P\mathrm{I}} \tag{3.1}$$

RJ = proportion of the time judges agree

$P\mathrm{I}$ = what agreement would we get by chance

In our study we have five judges and for each data set judges labeled 2145 sentences according to emotions. We can calculate agreement of judges from confusion matrix and below is our calculation which shows marked emotions' number by judges. This matrix's name is confusion matrix which is determined in chapter 4 at table 9. Following example shows kappa statistic calculation. Each cell of confusion matrix is filled with the number of judges who agreed that a belongs to an emotion' category which we determined. In the confusion matrix a=joy, b= fear, c=worry, d=sadness, e =anger, f= neutral.

$$\begin{bmatrix} & a & b & c & d & e & f & RJ \\ 1 & 5 & 0 & 0 & 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 1 & 0 & 0 & 4 & 0{,}6 \\ 3 & 0 & 1 & 2 & 1 & 0 & 1 & 0{,}1 \\ 4 & 0 & 0 & 1 & 0 & 1 & 3 & 0{,}3 \\ 5 & 0 & 0 & 1 & 0 & 0 & 4 & 0{,}6 \\ 6 & 0 & 0 & 1 & 0 & 0 & 4 & 0{,}6 \\ 7 & 0 & 2 & 3 & 0 & 0 & 0 & 0{,}6 \\ 8 & 0 & 1 & 4 & 0 & 0 & 0 & 0{,}4 \\ 9 & 0 & 1 & 4 & 0 & 0 & 0 & 0{,}6 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ 2145 & . & . & . & . & . & . & . \\ Total & 1060 & 501 & 2355 & 1106 & 981 & 4722 & 1395{,}9 \\ \mathrm{PI} & 0{,}00976 & 0{,}00218 & 0{,}04821 & 0{,}01060 & 0{,}00836 & 0{,}19384 & \end{bmatrix}$$

Hence;

N= 2145(number of sentences),

n = 5(number of judges),

k = 6(number of emotions),

Sum of cells = 2145*5= 10725

For column calculation,

$$P1 = \frac{5+0+0+\cdots}{2145*5} = \frac{1060}{2145*5} = 0{,}0988345$$

$$P2 = \frac{0+0+1+\cdots}{2145*5} = \frac{501}{2145*5} = 0{,}0467133$$

$$P3 = \frac{0+4+1+\cdots}{2145*5} = \frac{2355}{2145*5} = 0{,}2195804$$

$$P4 = \frac{0+0+1+\cdots}{2145*5} = \frac{1106}{2145*5} = 0{,}1031235$$

$$P5 = \frac{0+0+0+\cdots}{2145*5} = \frac{981}{2145*5} = 0{,}09146853$$

$$P6 = \frac{5+0+0+\cdots}{2145*5} = \frac{4722}{2145*5} = 0{,}44028$$

Total of columns; $\overline{\mathrm{PI}} = P1^2 + P2^2 + P3^2 + P4^2 + P5^2 + P6^2 = 0{,}273013139$.

For rows calculation, we did 2145 operations for each row. Rows calculation is doing below;

$$\mathrm{R1} = \frac{1}{5(5-1)}(5^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 - 5) = 1$$

$$\mathrm{R2} = \frac{1}{5(5-1)}(0^2 + 0^2 + 1^2 + 0^2 + 0^2 + 4^2 - 5) = 0{,}6$$

$$\mathrm{R3} = \frac{1}{5(5-1)}(0^2 + 1^2 + 2^2 + 1^2 + 0^2 + 1^2 - 5) = 0{,}1$$

$$\mathrm{RJ} = \mathrm{R1} + \mathrm{R2} + \mathrm{R3} + \cdots \mathrm{R2145} = 1395{,}9$$

$$\overline{RJ} = \frac{1}{((2145)(5)(5-1))}((1395{,}9)(5)(5-1)) = \frac{1395{,}9}{2145} = 0{,}65077$$

Finally we applied below formula;

$$K = \frac{\overline{RJ} - \overline{\mathrm{PI}}}{(1 - \overline{\mathrm{PI}})} = \frac{0{,}65077 - 0{,}273013139}{1 - 0{,}273013139} = 0{,}51961887$$

So, our value of k is 0,51961887 that means our result is moderate agreement.

3.2 Feature Types

3.2.1 Term frequency (tf) and inverse term frequency (idf)

In this thesis tf, idf and unigram used as feature types. Our data sets' instances are string so; we used StringToWordVector filters in WEKA with supervised discretization. StringToWordVector converts the properties represented by the knowledge of where it passes into a series of words in the dataset to the string properties. Using this filter we found which words were available in our text file, also calculated mean and StdDev of each word. These words are our features and variable of each words are weighted by the tf-idf all distinct words are considered as features which are unigram. Table 8 shows number of unigrams in our dataset.

Table 8 Unigram Count for Each Dataset

Data set	**#unigram**
Dataset 1	2769
Dataset 2	4097
Combined	6032

We used tf and idf to calculate frequency of words. tf score of a term is a value indicating the number of occurrences of the term in the document. Namely each document is represented by a vector of frequencies of the terms as d = (tf1; tf2;… tfm).

In many documents terms are used frequently, despite having no distinctive significance will have a high score. This problem can be solve with idf_tscore.

Theoretical term frequency (tf) means measure of term density in a document. Text is expressed with the frequency of terms contained in the text. Inverse document frequency (idf) means a measure of the informativeness of a term [37].

$$idf_t = \log \frac{N}{df_t} \tag{3.2}$$

N = total number of documents

df_t = number of documents containing term j

The numbers of occurrence of the terms are weighted as in the following formula unlike from tf which could be just a count of documents with the term.

$$\text{tf - i}df_{t,d} = tf_{t,d} \text{ x i}df_t \tag{3.3}$$

3.3 Weka

WEKA (Waikato Environment for Knowledge Analysis) is the most important and popular for Machine Learning applications. Weka is an open source data mining program which is developed on Java platform. It is improved by Waikato University in New Zealand. Figure 12 represents interface of WEKA [38]. It includes several well known machine learning algorithm as classifier used to train and test on different data set. Using Weka, can be done data mining processes, data analysis, visualize data sets and business intelligence applications.

Three kinds of data mining's operations can be done in WEKA. These operations are as classification, clustering and association. Also data pre-processing's transaction can be done. Several Machine Learning Classification Methods was applied on WEKA for our study. In our study we used methods of WEKA to analyze our data sets. These methods are NB, J48, and SVM. To do applications, we have to create format which can run in WEKA. C4.4, LibSVM's format, WEKA's format which is ARFF (Attribute Relationship File Format) and CSV are appropriate for this program. We used ARFF format, sample of ARFF represents in Figure 11.

```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
```

Figure 11 ARFF format for WEKA

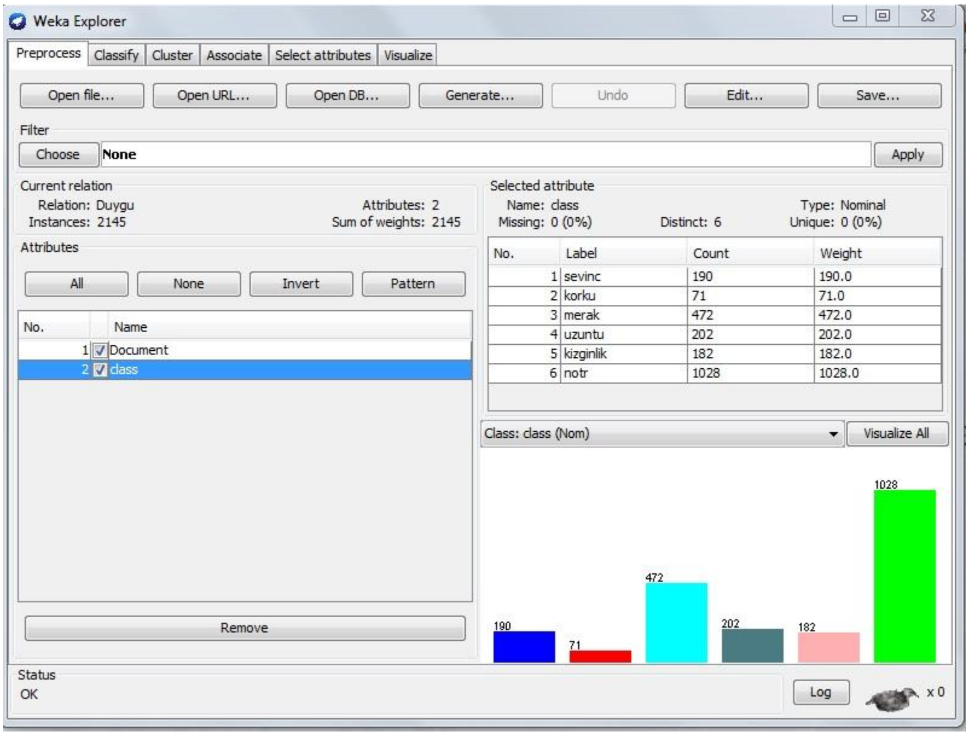

Figure 12 A Screenshot of WEKA

CHAPTER 4

EXPERIMENTAL RESULTS

In this chapter evaluation measures and experiment results are presented. In Section 4.1 evaluation measures are shown. Classification a result of first Dataset is given in Section 4.2 and classification results of second DataSet is given in Section 4.3. Combination of first Dataset and second DataSet Classification results are given in Section 4.3.

4.1 Evaluation Measures

Basic emotion classes are used in assessing the success of our study. Used measures are; Precision, Recall, F-measure. These measures increase the success of algorithms.

The success of the model depends on the number of instances assigned to the correct class and the number of instances assigned to the wrong class. We can represent the success of results by a confusion matrix as Table 9.

Table 9 Confusion Matrix

		Predicted Class	
		Class=1	Class= 0
Actual Class	Class=1	TP	FN
	Class= 0	FP	TN

TP = True – Positive

FN = False – Negative

FP = False –Positive

TN = True – Negative

In this matrix rows mean real number of test set's instances. Columns mean prediction class of the model.

4.1.1 Precision

Precision is the ratio of the number of correctly predicted items of a class in dataset divided by the number of total items that are predicted to be in that class. It is usually expressed as a percentage.

$$P = \frac{TP}{TP + FP} = \frac{|\text{correctly predicted items of a class}|}{|\text{number of total items(total predicted items of a class)}|}$$

4.1.2 Recall

Recall is represents the concept of how much of the items of a class is correctly predicted. It is usually expressed as a percentage.

$$R = \frac{TP}{TP + FN} = \frac{|\text{correctly predicted items of a class}|}{|\text{number of total items(true items of a class)}|}$$

4.1.3 F-measure

F-measure is the harmonic mean of precision and recall which is percentage of correct classification.

$$F = \frac{2\ x\ P\ x\ R}{P + R}$$

4.1.4 Accuracy

A result of accuracy represents the success of the classification algorithms to all instances.

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

4.2 Classification Results of DataSet 1

In the first Dataset four algorithms are applied separately to classify sentences with six emotions. This dataset includes 2145 instances. We extracted first Dataset's instances from Agatha CHRISTIE "Şampanyadaki Zehir" novel. All instances are dependent to each other because we created from sequential parts novel. Table 10 shows the number of all algorithms' correct and incorrect instances, also kappa statistics. For all methods' results obtained using 10-fold cross - validation which is explained in chapter 2.3. This study' data set divided into two sets as training and test which is determined in chapter 2.2.2. NB, J48 and SVM methods were applied to data set after the 10-fold cross - validation. In the KNN methods cross validate' variable is taken true and k-fold is taken 10.

The best results are obtained for the first Dataset when using SVM algorithm, which provides an accuracy of 84.5221%. This performance is also valid when the Kappa statistic is used, the result of SVM is best according to Kappa statistic. Successes of other algorithms are respectively; KNN, NB and J48.

Table 10 Algorithms' Results of First Dataset

Used Algorithms	Correctly Classified Instances	Incorrectly Classified Instances	Kappa Statistic	Accuracy
NB	1627	518	0.7093	75.8508 %
J48	1541	604	0.6615	71.8415 %
SVM	1813	332	0.8141	**84.5221 %**
KNN	1687	458	0.7429	78.648 %

Table 11 shows the detailed accuracy of NB method. All these result are taken using unigram and we obtained 2769 attributes. The best value of F-Measure is 0.821 for fear class, best value of recall is 0.982 again for fear class and best precision's value is 0.872 for joy class for NB. In NB, F-measure value is lover in neutral class according to other methods.

Table 11 Detailed Accuracy of NB for First Dataset

	TP	FP	Precision	Recall	F-Measure	Class
NB	0.734	0.020	0.872	0.734	0.797	Joy
	0.982	0.088	0.706	0.982	0.821	Fear
	0.723	0.038	0.779	0.723	0.750	Worry
	0.833	0.047	0.778	0.833	0.805	Sadness
	0.882	0.067	0.746	0.882	0.808	Anger
	0.356	0.031	0.689	0.356	0.470	Neutral
Weighted Avg.	0.759	0.050	0.760	0.759	0.744	

Table 12 shows the detailed accuracy of SVM method. The best value of F-Measure is 0.982 for fear class, best value of recall is 0.992 again for fear class and best precision's value is 0.972 for fear class for SVM.

Table 12 Detailed Accuracy of SVM for First Dataset

	TP	FP	Precision	Recall	F-Measure	Class
SVM	0.839	0.034	0.822	0.839	0.830	Joy
	0.992	0.006	0.972	0.992	0.982	Fear
	0.807	0.034	0.816	0.807	0.811	Worry
	0.853	0.028	0.858	0.853	0.856	Sadness
	0.921	0.28	0.878	0.921	0.899	Anger
	0.635	0.055	0.691	0.635	0.662	Neutral
Weighted Avg.	0.845	0.030	0.842	0.845	0.844	

Table 13 shows the detailed accuracy of J48 method. The best value of F-Measure is 0.980 for fear class, best value of recall is 0.976 again for fear class and best precision's value is 0.984for fear class for J48. In J48, F-measure value is lover in neutral class according to other methods.

Table 13 Detailed Accuracy of J48 for First Dataset

	TP	FP	Precision	Recall	F-Measure	Class
J48	0.648	0.027	0.816	0.648	0.722	Joy
	0.976	0.003	0.984	0.976	0.980	Fear
	0.610	0.028	0.804	0.610	0.694	Worry
	0.774	0.116	0.568	0.774	0.656	Sadness
	0.803	0.063	0.740	0.803	0.770	Anger
	0.457	0.101	0.468	0.457	0.462	Neutral
Weighted Avg.	0.718	0.056	0.733	0.718	0.719	

Table 14 shows the detailed accuracy of KNN method. The best value of F-Measure is 0.938 for fear class, best value of recall is 0.992 again for fear class and best precision's value is 0.890 for fear class for KNN.

Table 14 Detailed Accuracy of KNN for First Dataset

	TP	FP	Precision	Recall	F-Measure	Class
KNN	0.839	0.041	0.789	0.839	0.813	Joy
	0.992	0.027	0.890	0.992	0.938	Fear
	0.717	0.019	0.873	0.717	0.788	Worry
	0.833	0.036	0.819	0.833	0.826	Sadness
	0.959	0.123	0.635	0.959	0.764	Anger
	0.336	0.012	0.848	0.336	0.481	Neutral
Weighted Avg.	0.786	0.044	0.807	0.786	0.771	

4.3 Classification Results of DataSet 2

This dataset includes 2145 instances as a first DataSet. In table 15 is representing percentages of all algorithms' correct and incorrect instances, also kappa statistics. As the DataSet 1, 10-fold cross – validation used for all methods in DataSet 2. NB, J48 and SVM methods were applied to data set after the 10-fold cross - validation. In the KNN methods cross validate' variable is taken true and k-fold is taken 10. A best algorithm for second DataSet is SVM's result which is 83.9161 %. As shown of kappa statistic's result of SVM is best according to other results of algorithms like a first Dataset and. Success of other algorithms respectively; KNN, NB and J48.

Table 15 Results Algorithms' Results of Second Dataset

Used Algorithms	**Correctly Classified Instances**	**Incorrectly Classified Instances**	**Kappa Statistic**	**Accuracy**
NB	1594	551	0.6904	74.3124 %
J48	1452	693	0.6107	67.6923 %
SVM	1800	345	0.8068	**83.9161 %**
KNN	1719	426	0.7615	80.1399 %

Table 16 shows the detailed accuracy of NB methods. All these result are taken using unigram and we obtained 2769 attributes. The best value of F-Measure is 0.857 for fear class, best value of recall is 0.982 again for fear class and best precision's value is 0.877 for sadness class for NB.

Table 16 Detailed Accuracy of NB for Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
NB	0.728	0.023	0.856	0.728	0.787	Joy
	0.982	0.067	0.761	0.982	0.857	Fear
	0.616	0.029	0.799	0.616	0.696	Worry
	0.763	0.021	0.877	0.763	0.816	Sadness
	0.908	0.135	0.599	0.908	0.722	Anger
	0.414	0.036	0.689	0.414	0.517	Neutral
Weighted Avg.	0.743	0.054	0.760	0.743	0.734	

Table 17 shows the detailed accuracy of SVM method. The best value of F-Measure is 0.978 for fear class, best value of recall is 0.982 again for fear class and best precision's value is 0.974 for fear class for SVM.

Table 17 Detailed Accuracy of SVM for Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
SVM	0.869	0.030	0.841	0.869	0.855	Joy
	0.982	0.006	0.974	0.982	0.978	Fear
	0.726	0.032	0.811	0.726	0.766	Worry
	0.828	0.027	0.857	0.828	0.842	Sadness
	0.887	0.26	0.885	0.887	0.886	Anger
	0.721	0.072	0.661	0.721	0.690	Neutral
Weighted Avg.	0.839	0.032	0.841	0.839	0.840	

Table 18 shows the detailed accuracy of J48 method. The best value of F-Measure is 0.960 for fear class, best value of recall is 0.971 again for fear class and best precision's value is 0.949 for fear class for J48. In J48, F-measure value is lover in neutral class according to other methods.

Table 18 Detailed Accuracy of J48 for Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
J48	0.707	0.026	0.835	0.707	0.766	Joy
	0.971	0.011	0.949	0.971	0.960	Fear
	0.512	0.050	0.656	0.512	0.575	Worry
	0.624	0.029	0.810	0.624	0.705	Sadness
	0.841	0.198	0.485	0.841	0.615	Anger
	0.353	0.076	0.475	0.353	0.405	Neutral
Weighted Avg.	0.677	0.067	0.701	0.677	0.675	

Table 19 shows the detailed accuracy of KNN method. The best value of F-Measure is 0.945 for fear class, best value of recall is 0.982 again for fear class and best precision's value is 0.947 for sadness class for KNN.

Table 19 Detailed Accuracy of J48 for Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
KNN	0.857	0.087	0.646	0.857	0.737	Joy
	0.982	0.021	0.910	0.982	0.945	Fear
	0.616	0.011	0.916	0.616	0.737	Worry
	0.802	0.009	0.947	0.802	0.869	Sadness
	0.874	0.024	0.890	0.874	0.882	Anger
	0.647	0.086	0.592	0.647	0.618	Neutral
Weighted Avg.	0.801	0.039	0.821	0.801	0.803	

4.3 Combination of First and Second Datasets Classification Results

We combined two data sets, yielding 4290 instances totally. Table 20 shows the number of all algorithms' correct and incorrect instances, also kappa statistics. A best algorithm for first Dataset and second DataSet is SVM's result which is **82.5175** %. Success of other algorithms respectively; NB, KNN and J48.

Table 20 Results Algorithms' results First Dataset and Second Dataset

Used Algorithms	**Correctly Classified Instances**	**Incorrectly Classified Instances**	**Kappa Statistic**	**Accuracy**
NB	3331	959	0.7317	77.6457 %
SVM	3540	750	0.7902	**82.5175 %**
KNN	3202	1088	0.6957	74.6387 %
J48	2997	1293	0.6383	69.8601 %

Table 21 shows the detailed accuracy of NB method. All these result are taken using unigram and we obtained 2769 attributes. The best value of F-Measure is 0.895 for fear class, best value of recall is 0.982 again for fear class and best precision's value is 0.826 for fear class for NB.

Table 21 Detailed Accuracy of NB for First Dataset and Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
NB	0.813	0.043	0.789	0.813	0.801	Joy
	0.982	0.042	0.826	0.982	0.895	Fear
	0.789	0.046	0.774	0.789	0.781	Worry
	0.786	0.043	0.785	0.786	0.785	Sadness
	0.815	0.047	0.776	0.815	0.795	Anger
	0.474	0.046	0.671	0.474	0.556	Neutral
Weighted Avg.	0.776	0.045	0.770	0.776	0.769	

Table 22 shows the detailed accuracy of SVM method. The best value of F-Measure is 0.979 for fear class, best value of recall is 0.again for fear class and best precision’s value is 0.967 for fear class for SVM.

Table 22 Detailed Accuracy of SVM for First Dataset and Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
SVM	0.853	0.040	0.809	0.853	0.830	Joy
	0.992	0.007	0.967	0.992	0.979	Fear
	0.786	0.030	0.840	0.786	0.812	Worry
	0.838	0.029	0.853	0.838	0.845	Sadness
	0.877	0.043	0.805	0.877	0.839	Anger
	0.606	0.062	0.663	0.606	0.633	Neutral
Weighted Avg.	0.825	0.035	0.8823	0.825	0.823	

Table 23 shows the detailed accuracy of KNN method. The best value of F-Measure is 0.916 for fear class, best value of recall is 0.992 again for fear class and best precision’s value is 0.917 for joy class for KNN. In KNN, F-measure value is lover in neutral class according to other methods.

Table 23 Detailed Accuracy of KNN for First Dataset and Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
KNN	0.807	0.015	0.917	0.807	0.859	Joy
	0.992	0.035	0.851	0.992	0.916	Fear
	0.643	0.013	0.906	0.643	0.752	Worry
	0.807	0.032	0.833	0.807	0.820	Sadness
	0.943	0.200	0.486	0.943	0.641	Anger
	0.287	0.010	0.858	0.287	0.430	Neutral
Weighted Avg.	0.746	0.051	0.808	0.746	0.736	

Table 24 shows the detailed accuracy of KNN method. The best value of F-Measure is 0.960 for fear class, best value of recall is 0.966 again for fear class and best precision's value is 0.954 for worry class for KNN.

Table 24 Detailed Accuracy of J48 for First Dataset and Second Dataset

	TP	FP	Precision	Recall	F-Measure	Class
J48	0.666	0.025	0.841	0.666	0.743	Joy
	0.966	0.009	0.954	0.966	0.960	Fear
	0.649	0.035	0.788	0.649	0.712	Worry
	0.632	0.032	0.797	0.632	0.705	Sadness
	0.785	0.155	0.503	0.785	0.613	Anger
	0.494	0.010	0.484	0.494	0.489	Neutral
Weighted Avg.	0.699	0.060	0.728	0.699	0.704	

CHAPTER 5

CONCLUSION

We obtained good results using the different authors' books with ML methods. In this thesis, we studied emotion detection on novels. Turkish Novels weren't used before as a data set in such as similar kinds of studies. To decide to use novel, other studies have investigated to learn which resources used as a data set. Agatha CHRISTIE "Şampanyadaki Zehir" and Zülfü LİVANELİ "Serenad" authors' achievements were used in this thesis. Two data sets are analyzed and compared and totally 2145 instances available for each data set. Both data sets handled using six basic emotions by judges. The purpose of this study was to obtain successful results using the methods of classification with emotion detection for sentences of novels. In addition a decrease wasn't observed on performance of classification methods in sentences which obtained from two different novels.

Applied methods is done by ML approaches which try to classify the test data after learning process carried out using the training data, therefore such methods are supervised ML processes as our study. The results obtained in machine learning approaches have been formed by 10-fold cross validation. Used classifiers methods are J48, KNN, NB and SVM. Feature types are tf, idf and unigram.84.5221 % is for first Dataset, 83.9161 % second DataSet and 82.5175 % for combined data's result. These two data sets were generated by different authors' book and same instances were used for both data sets also same methods were applied to both data sets. In Z. Boynukalın' study CNB gave best result which is 76.83% with five classes for ISEAR dataset using 10 fold cross validation.

In the future more emotions can be defined to analyze specific emotions using more seed words to understand novel's emotion. Either, comprehensive adjectives library may be generated that affect emotions, using this library we can assign these adjectives to emotion categories and can be done emotion detection to our dataset. For all that can be analyzed impact of adjectives to emotions. In addition data set may be analyzed according infection and suffix to Turkish syntax, after that we can apply ML processes to look increase or not.

REFERENCES

1. **Shivhare S. N., (2012),** *"Emotion Detection From Text",* Computer Science & Information Technology (CS & IT), pp. 371–377.

2. **Sebea N. Cohenb I., Geversa T., and Huangc T. S.**, "*Multimodal Approaches for Emotion Recognition: A Survey*", USA, vol.16, no.2.

3. **P. Ekman, M. J. Power, (1999.),** *"In Handbook of Cognition and Emotion"*, pp. 45-60.

4. **Kleinginna P. R., & Kleinginna A. M., (1981),** *"A Categorized List of Emotion Definitions, with Suggestions for a Consensual Definition. Motivation and Emotion"*, vol.5, no.4, pp. 345–379.

5. **Kim S. M., (2011),** *"Recognising Emotions and Sentiments in Text"*,April.

6. **Russell J. A., (1979),** *"Affective Space is Bipolar. Journal of Personality and Social Psychology",* vol.37, no.3, pp.345-356.

7. **Russell J. A., (1980),** *"A Circumplex Model of Affect. Journal of Personality and Social Psychology",* vol.39, no.6, pp.1161-1178.

8. **Russell J. A., Lewicka, M., & Niit, T., (1989),** *"A Cross-Cultural Study of a Circumplex Model of Affect. Journal of Personality and Social Psychology"*, vol.57, no.5, pp.848-856.

9. **Plutchik R., (1980),** *"Emotion: A Psychoevolutionary Synthesis: Harpercollins College Div",* New York.

10. **Pang B. and Lee, L., (2005),** *"Seeing stars: Exploiting Class Relationships for Sentiment Categorization with Rspect to Rating Scales. In Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL'05)"*, Ann Arbor, Michigan, USA, pp.115-124.

11. **Dasu T. ve Johnson, T., (2003),** "*Explarotary Data Mining and Data Cleaning*", pp 240–251.

12. **Bruce A. Wooley, (1999),** "*Scaling Clustering for the Data Mining Step in Knowledge Discovery*", Mississippi State University, MS 39762, pp.2-3.

13. **Wolpert David H., (1996),** *"The Lack of a Priori Distinctions Between Learning Algorithms. Neural Computation"*, vol.8, no.7, pp.1341–1390.

14. **S. B. Kotsiantis, (2007),** "*Supervised Machine Learning: A Review of Classification Techniques*", vol.31, pp.249–268.

15. **McCallum A., Nigam K., (2003),** *"A Comparison of Event Models for Naive Bayes Text Classification",* Journal of Machine Learning Research, vol.3, pp.1265-1287.

16. **J. D. M. Rennie, L. Shih, J. Teevan, and D. R. Karger, (2003),** *"Tackling the poor assumptions of naive bayes text classification, ICML"*, pp.616-623.

17. **C. Cortes and V. Vapnik, (1995),** *"Support-Vector Networks. Machine Learning"*, vol.20, pp.273-297.

18. **J. R. Quinlan, (1985),** "*Induction of Decision Trees, Machine Learning*", vol.1, no.1, pp.81-106.

19. **Witten I. H., & Frank E., (2005),** *"Data Mining: Practical Machine Learning Tools and Techniques",* 2nd Edition, Morgan Kaufmann, San Francisco, pp.429-451.

20. **P. Hall, B. U. Park, R. J. Samworth, (2008),** "*Choice of Neighbor Order in Nearest-Neighbor classification",* vol.36, pp.2135-2152.

21. **Esuli A., Sebastiani F., (2006),** *"SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining"*, Proceedings of LREC, pp.417-422.

22. **Paltoglou G., Thelwall M., (2010),** *"A Study of Information Retrieval Weighting Schemes for Sentiment Analysis",* Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, pp.1386-1395.

23. **Aman S. ,(2007),** *"Recognizing Emotions in Text",* University of Ottawa, Master's Thesis, pp.9-20.

24. **Go A., Bhayani, R., and Huang, L., (2009)**, *“Twitter Sentiment Classification Using Distant Supervision”,* CS224N Project Report, Stanford University Department of Computer Science, CA, US, pp.1-2.

25. **Cetin M. and Amasyalı, M. F., (2013)** “*Supervised and Traditional Term Weighting Methods for Sentiment Analysis”,* 21. Sinyal İşleme ve İletişim Uygulamaları, Uluslararası Kıbrıs Universitesi, Girne, KKTC, pp.1-4.

26. **Akbaş E., (2012),** *“Aspect Based Opinion Mining on Turkish Tweets”,* M. Sc. Thesis,Bilkent University Graduate School of Natural and Applied Sciences, Master’s Thesis.

27. **Pang Bo, Lillian L., and Shivakumar V., (2002),** *"Thumbs up?: Sentiment Classification Using Machine Learning Techniques."* Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, vol.10, pp.79-86.

28. **Bradley M. M., Lang, P. J., Bradley, M. M., and Lang, P. J., (1999),** *“Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings”,* Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, US.

29. **C. Alm and D. Roth and R. Sproat, (2005),** “*Emotions From Text: Machine Learning for Text-Based Emotion Prediction*”. EMNLP, pp.579-586.

30. **Sevindi B. İ., (2013),** “*Comparision of Supervised of and Dictionary Based Sentiment Analysis Approaches on Turkish Text*”, Master’s Thesis.

31. **Boynukalın Z., (2012),** *“Emotion Analysis of Turkish Texts by Using Machine Learning Methods”,* M. Sc. Thesis, Middle East Technical Unıversity Graduate School of Natural and Applied Sciences.

32. **K.R. Scherer and H. G. Wallbott, (1994),** *“Evidence for Universality and Cultural Variation of Differential Emotion Response Patterning”*, Journal of Personality and Social Psychology, vol.66, pp.310-328.

33. **Parrott W. G., (2001),** “*Emotions in Social Psychology: Essential Readings*”,Psychology Press, UK.

34. **Dodds P. S., Harris K. D., Kloumann I. M., Bliss C. A., and Danforth C. M., (2011*),"***Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter"*, PloS One, vol.6, no.12, pp.26752.

35. **Şimşek, M. U., (2012),** *"Sosyal Ağlarda Veri Madenciliği Üzerine Bir Uygulama",* Yuksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

36. **Manning C. D., P. Raghavan, and H. Schutze, (2008), "***Introduction to Information Retrieval***",** Cambridge Univ. Press.

37. **Justin M. and Tim F., (2009),** "*Delta TFIDF: An Improved Feature Space for Sentiment Analysis*", San Jose, California, USA. The AAAI Press.

38. **Ian H. Witten; Eibe F., Len T., Mark H., Geoffrey H., and Sally J., C., (1999),** *"Weka: Practical Machine Learning Tools and Techniques with Java Implementations"*. Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems,vol.06, no.26, pp.192–196.

APPENDICES A

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Başdemir, Bilge

Date and Place of Birth: 02 July 1985, Ankara

Marital Status: Single

Phone: 0533 742 00 48

Email: bilgebasdemir_@hotmail.com

EDUCATION

Degree	**Institution**	**Year of Graduation**
M.Sc.	Çankaya University, Computer Engineering	2015
B.Sc.	Girne American University, Computer Engineering	2010
High School	50. Yıl Super High School	2003

WORK EXPERIENCE

Year	**Place**	**Enrollment**
2012- Present	Digitest Electronic	Project Management
2010-2012	Girne American University	Social Media Network Control & Department of Digital Marketting
04.2012-10.2012	Go-Tasarım	Web Interface Designer
2008 July	T.C Başbakanlık Müsteşarlığı	Intern

2005-2007	Normatif (Monthly Journal of Business and Economics)	Decipher
2004	ÖSYM	Part-Time Staff

FOREIN LANGUAGES

Upper-intermediate English

HOBBIES

Reading, Travel, Listening Music, Theater