

T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**TWITTER ÜZERİNDEKİ TÜRKÇE MESAJLARDA
VERİ MADENCİLİĞİYLE DUYGU ANALİZİ**
YÜKSEK LİSANS TEZİ

TEZİ HAZIRLAYAN:
BURÇİN ADAK KAPLAN

İstanbul, 2016

T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI

**TWITTER ÜZERİNDEKİ TÜRKÇE MESAJLARDA
VERİ MADENCİLİĞİYLE DUYGU ANALİZİ**
YÜKSEK LİSANS TEZİ

TEZİ HAZIRLAYAN:

BURÇİN ADAK KAPLAN

ÖRENCİ NO:

130820021

DANIŞMAN:

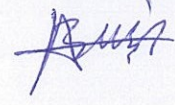
YARDIMCI DOÇENT DOKTOR EDİZ ŞAYKOL

İstanbul, 2016

YEMİN METNİ

Yüksek lisans tezi olarak sunduğum “TWITTER ÜZERİNDEKİ TÜRKÇE MESAJLARDA VERİ MADENCİLİĞİYLE DUYGU ANALİZİ” başlıklı bu çalışmanın, bilimsel ahlak ve geleneklere uygun şekilde tarafımdan yazıldığını, yararlandığım eserlerin tamamının kaynaklarda gösterildiğini ve çalışmamın içinde kullandıkları her yerde bunlara atıf yapıldığını belirtir ve bunu onurumla doğrularım. 23/01/2017

Burçin Adak Kaplan



T.C.
BEYKENT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZ SAVUNMA SINAVI SONUÇ TUTANAĞI

Beykent Üniversitesi
Fen Bilimleri Enstitüsü Müdürlüğü'ne,

BURGIN ADAK KAPLAN

Aşağıda tez adı belirtilen yüksek lisans öğrencisi.....no'lu.....'in 19.1.17 tarihinde yapılan tez savunma sınavı¹ sonucunda 45 dakika süreyle sunduğu ve savunduğu tezi hakkında² oybirliğiyle KABUL kararı verilmiştir.

Bilgilerinize saygılarımızla arz ederiz.

Anabilim Dalı :
Programı :
Tez Başlığı³ :

Tez Sınav Jürisi

Öğretim Üyesi

Danışman

: Yrd. Doç. Dr. Ediz SAYGIL

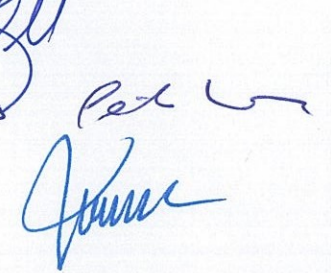
Üye

: Doç. Dr. Gökhan SİLİHTAROĞLU

Üye

: Yrd. Doç. Dr. Tuncay KARAGÜLER

İmza



¹ Jüri üyeleri söz konusu tezin kendilerine teslim edildiği tarihten itibaren en geç bir ay içinde toplanarak öğrenciyi tez savunma sınavına alır. Belirlenen günde yapılamayan jüri toplantısı, katılanların hazırladığı bir tutanakla enstitü yönetimine bildirilir. Bu durumda jüri en geç onbeş gün içinde toplanarak adayı tez savunma sınavına alır. Tez savunma sınav süresi en az 45 dakikadır. Yüksek lisans tez savunma sınavı, tez çalışmasının sunulması ve bunu izleyen soru-yanıt bölümlerinden oluşur ve dinleyiciye açıktır. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-3)

² Tez sınavının tamamlanmasından sonra jüri, tez hakkında “kabul”, “düzeltme” veya “red” kararı verir. Jüri başkanı, jüri üyelerince imzalanmış sınav tutanağını, tez sınavını izleyen üç gün içinde ilgili enstitü yönetimine teslim eder. Tezi başarısız bulunan öğrencinin Enstitü ile ilişkisi kesilir. Tezi hakkında düzeltme kararı verilen öğrenci en geç üç ay içinde gerekli düzeltmeleri yaparak ve yönetmelikte belirtilen usullere uygun olarak tezini aynı jüri önünde yeniden savunur. Bu savunma sınavında da tezi kabul edilmeyen öğrencinin enstitü ile ilişkisi kesilir. (Beykent Lisansüstü eğitim ve Öğretim Yönetmeliği-Madde30-4)

³ İleride doğabilecek saklıkların engellenmesi için tezin başlığını yazılması gerekmektedir.

Adı ve Soyadı : Burçin Adak Kaplan
Danışmanı : Yardımcı Doçent Ediz Şaykol
Türü ve Tarihi : Yüksek Lisans, 2016
Alanı : Bilgisayar Mühendisliği

Anahtar Kelimeler : Semantik, Anlambilim, Duygu Analizi, Veri madenciliği

ÖZET

TWITTER ÜZERİNDEN GÖNDERİLEN TÜRKÇE MESAJLARDA VERİ MADENCİLİĞİ TEKNİKLERİ KULLANILARAK DUYGU ANALİZİ YAPILMASI

Bu çalışmada, yazılı Türkçe metinlerde duygu analizi yapılabilmesi amacıyla yapılan semantik çalışmalara katkı sağlanması hedeflenmiştir. Ayrıca, Türkçe metinlerde duygu analizinin yapılabilmesi ile Türkçe dilinde konuşan, ortak zevk, düşünce ve duygulara sahip insanların bir araya getirilebileceği de öngörülmüştür. Bu sayede, internet üzerindeki sosyal paylaşım siteleri de, daha başarılı yapay zekâ yapıları üzerine kurgulayabileceklerdir. Çalışma kapsamında, sosyal medya ağı olan Twitter kullanıcılarının tivitleri analiz edilmiştir. Paylaşılan tivitlerin ifade ettiği duygular dört farklı sınıfta sınıflandırılmıştır. Bu sınıflar 'Mutluluk', 'Kızgınlık', 'Üzüntü' ve 'Şaşkınlık' gruplarında toplanmıştır. Tivitlerin sınıflandırılmasının yapılabilmesi için tivitler toplandıktan sonra yazım hataları olması ihtimaline karşılık bütün içerikler 'Zemberek' kütüphanesi yardımıyla yazım hatalarından arındırılmıştır. Yazımları düzeltilen tivitler gönüllüler tarafından dört grupta etiketlenmiştir. Çalışma Twitter'dan toplanan tivitleri karar ağacı ve bulanık kural öğrenme teknikleri kullanılarak analiz edilmiş ve elde edilen sonuçlar incelenmiştir.

Name and Surname : Burçin Adak Kaplan
Supervisor : Yardımcı Doçent Ediz Şaykol
Degree and Date : Master, 2016
Major : Computer Engineering

Key Words : Semantic, Sentiment, Emotion Analysis, Data mining

ABSTRACT

SENTIMENT ANALYSIS ON TURKISH TWITTER MESSAGES BY USING DATA MINING

One of the aims of this thesis is make a contribution on semantic works for sentiment analysis on Turkish texts. Also, it is called for people who talk in Turkish and having common ground can be brought close together by achieving sentiment analysis in Turkish texts. Thus, social networking sites can be built based on more successful artificial intelligence systems.

Tivits of users in a social media, 'Twitter', were analysed as in scope of this thesis. Emotions stated by shared tivits were classified under four main categories. These categories are 'Happiness', 'Anger', 'Sadness', and 'Confusion'.

All typo mistakes of tivits were proofread with the help of 'Zemberek' library for classifying these accurately. Proofread tivits were labeled on these four categories by volunteers. After this, these tivits and related results were examined by using the techniques of 'Decision tree' and 'Fuzzy Rules'.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
İÇİNDEKİLER.....	iii
TABLolar LİSTESİ	v
ŞEKİLLER LİSTESİ.....	vii
KISALTMALAR.....	x
1. GİRİŞ	1
2. BENZER ÇALIŞMALAR.....	4
3. KAVRAMLAR.....	11
3.1 Duygu Analizi	11
3.2 Twitter	11
3.3 Uygulama programlama ara yüzü (API)	13
3.4 Twitter API.....	14
3.5 Oauth.....	15
3.6 JavaScript Nesne Notasyonu (JSON)	16
3.7 Linq2Twitter.....	17
3.8 NoSQL.....	17
3.9 DocumentDB.....	18
3.10 Zemberek Kütüphanesi	18
3.11 Knime	19
3.12 Karar Ağacı	20
3.13 Bulanık Mantık.....	21
3.14 C# programlama dili.....	22

3.15 Veri Madenciliđi	23
4. YÖNTEM	24
4.1 Verilerin Toplanması	25
4.2 Verilerin Temizlenmesi	29
4.3 Verilerin Hazırlanması	35
4.4 Verilerin Knime Programı ile Analiz Edilmesi	38
5. SONUÇ	54
6. DEĞERLENDİRME	55
KAYNAKLAR	58
ÖZGEÇMİŞ	60

TABLolar LİSTESİ

Tablo	1.	Örnek	Çekilen
Tivitler.....		28	
Tablo	2.	Örnek	Temizlenen
Tivitler.....		34	
Tablo 3.	Örnek Etiketlenen Tivitler		36
Tablo 4.	Karar Ağacı ile Duygu Analizi Sonuçları		54
Tablo 5.	Fuzzy Rule Learner ile Duygu Analizi Sonuçları		54
Tablo.6.	Analiz edildiğinde hatalı dönen örnek duygu ifadeleri		56

ŞEKİLLER LİSTESİ

Şekil.1. Botego tarafından yayımlanan örnek duygu analizi verileri.....	7
Şekil.2. API Mantığı.....	14
Şekil.3. Knime Örnek Ekran Görüntüsü	19
Şekil.4. Karar Ağacı Örneği	20
Şekil.5. Bulanık Mantık	21
Şekil.6. Duygu Analizi Programı Aşamaları.....	24
Şekil.7. Veri Çekme Modülü Kodlar Bölüm 1.....	26
Şekil.8. Veri Çekme Modülü Kodlar Bölüm 2.....	27
Şekil.9. Veri Temizleme Modülü Kodlar.....	30
Şekil.10. Knime platformunda verilerin analizi için hazırlanan program.....	38
Şekil.11. Knime programında kullanılan File Reader modülü konfigürasyonu	39
Şekil.12. Knime programında kullanılan String To Document modülü konfigürasyonu.....	39
Şekil.13. Knime programında kullanılan Column Filter modülü konfigürasyonu.....	40
Şekil.14. Knime programında kullanılan Punctuation Erasure modülü konfigürasyonu.....	41
Şekil.15. Knime programında kullanılan N Chars Filter modülü konfigürasyonu.....	42
Şekil.16. Knime programında kullanılan Case Converter modülü konfigürasyonu.....	43
Şekil.17. Knime programında kullanılan Snowball Stemmer modülü konfigürasyonu.....	44
Şekil.18. Knime programında kullanılan Bag Of Words Creator modülü konfigürasyonu.....	45
Şekil.19. Knime programında kullanılan Group By modülü konfigürasyonu	46

Şekil.20. Knime programında kullanılan Row Filter modülü konfigürasyonu	47
Şekil.21. Knime programında kullanılan Reference Row Filter modülü konfigürasyonu	48
Şekil.22. Knime programında kullanılan Document Vector modülü konfigürasyonu	49
Şekil.23. Knime programında kullanılan Partitioning modülü konfigürasyonu	50
Şekil.24. Knime programında kullanılan Decision Tree Learner modülü konfigürasyonu	51
Şekil.25. Knime programında kullanılan Fuzzy Rule Learner modülü konfigürasyonu	52
Şekil.26. Knime programında kullanılan Scorer modülü konfigürasyonu....	53

KISALTMALAR

API: Uygulama Programlama Arayüzü (Application programming interface)

Tivit: Twitter, kullanıcıların paylaştıkları 140 karakterlik mesaj.

JSON: JavaScript obje notasyonu



1. GİRİŞ

Günümüzde İnternet teknolojilerinin gelişmesine bağlı olarak insanların hayatları daha da kolaylaşmıştır. İnternetin sağladığı önemli unsurlardan olan mesafelerin ortadan kaldırılması sayesinde insanlar o anda buldukları yerden kilometrelerce uzaktaki insanlarla görüşebilmektedirler. Ayrıca, kullanıcılar internet ortamında diğer insanlarla günümüzde Facebook, Foursquare, Twitter gibi sosyal platformlardan iletişim kurmakta, o sırada yaşamış oldukları bir olayı paylaşabilmekte veya bir olay hakkındaki duygu ve düşüncelerini dile getirebilmektedirler. Bu iletişimlerini de sesli, görüntülü ve/veya yazılı olarak yapabilmektedirler. Ancak eğer görüntülü veya sesli bir iletişim kurulmaktaysa duygular kolayca ifade edilebilirken yazılı iletişimlerde duygular tam olarak ifade edilemeye bilmektedir.

İnternet üzerinde yapılan yazışmalarda kullanıcılar karşısındakilere duygularını anlatabilmek için ekstra açıklamalar yazmaya ya da emoji denilen ve duygu belirten ikonlar kullanmaya gereksinim duymaktadırlar. Yazılan kelimelerden ve metinlerden duygu çıkarımına yönelik çalışmalar ise Türkçe dili için henüz istenilen aşamaya gelememiştir.

Çalışmada Türkçe dilindeki yazılı kelime ve kelime gruplarından duygu analizi yapılabilmesi hedeflenmiştir. Bu amaçla insanların Twitter sosyal ağı üzerinde Türkçe dilinde paylaştıkları tivitler ele alınmıştır. Çalışmayla paylaşılan Türkçe tivitlerde kişilerin aslen hangi duyguyu yansıttığının bilgisayarlar tarafından anlaşılabilmesi amaçlanmıştır.

Günümüzde Twitter artık yalnızca bir sosyal paylaşım ve mikroblog sitesi olmasından çok daha fazla anlam ifade etmektedir. Dünya üzerinde kullanıcılar hem ilgi alanları hakkında bilgi sahibi olabilmek, hem istedikleri konuma göre gündemi ilk ağızdan öğrenebilmek hem de sevinç, şikayet, kınama gibi hislerini paylaşmak için Twitter'ı kullanmaktadırlar.

Twitter, insanların sevinç, şikayet, kınama gibi duygularını paylaştıkları platform olma özelliği sayesinde ve anlık olarak farklı farklı kullanıcılardan yüksek

rakamlarda tivitler paylaşıp çeşitlilik sağlaması sebebiyle metinden duygu analizi yapılabilmesi için en doğru platform olarak görünmektedir. Bu sebeple de çalışmada twitter sosyal ağı üzerinden paylaşılan tivitlerin kullanılmasına karar verilmiştir.

Çalışmanın ilk aşamasında Twitter’da Türkçe olarak paylaşılan tivitler Twitter’in sağladığı API aracılığıyla toplanmıştır. Toplanan büyük miktarda verinin saklanıp, çok hızlı bir şekilde işlenebilmesi için bir NoSQL veri tabanı olan DocumentDB kullanılmıştır.

Twitter’ın özelliği olarak kullanıcıların en fazla 140 karakterden oluşan paylaşımları olan tivitleri gönderilip okunabilmektedir. Tivitlerin 140 karakterle sınırlı olması kullanıcıların istediklerini ifade ederken çok sayıda kısaltma ve yazım hataları yapmasına yol açmaktadır. Çalışmanın sağlıklı bir şekilde çalışması için bu yazım hataları ve kısaltmaların düzeltilmesi amacıyla Zemberek kütüphanesinden faydalanılmıştır. API aracılığıyla toplanan veriler veri tabanına kaydedilmeden önce yazım hatalarından bu kütüphane aracılığıyla arındırılmıştır.

Daha sonra, Tivitlerin ifade ettiği duygular ‘Mutluluk’, ‘Kızgınlık’, ‘Üzüntü’ ve ‘Şaşkınlık’ olmak üzere dört farklı sınıfta gönüllü kullanıcılar yardımıyla etiketlenmiştir. 100 farklı gönüllü, kişi başı ortalama 1500 adet tivit her birinin duygu sınıflandırmasını yapmıştır. Bu şekilde toplam 62347 adet tivit her birinin ifade ettiği duygular toplanmıştır. Bir tivitte en az 2 kullanıcı aynı duyguyu etiketlemişse bu tivit o duyguyu ifade ettiği kabul edilmiştir. Her bir tivit için aynı zamanda gönüllüler birden fazla duygu da seçebilmişlerdir.

Elde edilen veri kümesi veri analizi programlarından Knime kullanılarak 70’e 30 oranında öğrenme ve tahmin gruplarına ayrılmıştır. Öğrenme grubu karar ağacı ve bulanık kural yöntemleri kullanılarak analiz edilmiştir. Analiz sonuçları kullanılarak kalan %30’luk kısım tahmin edilmiştir.

Çalışmanın ikinci bölümünde benzer çalışmalar incelenip, mevcut çalışma ile karşılaştırmaları yapılmıştır. Üçüncü bölümde çalışmanın yöntemi, kullanılan araç ve kavramlar detaylı bir şekilde anlatılmıştır. Çalışmanın dördüncü bölümünde ise çalışma sonucunda elde edilen bulgular karşılaştırmalı olarak incelenip

yorumlanmıřtır. Sonu ve gelecekte yapılması gerekenler ise beřinci blmde incelenmiřtir.



2. BENZER ÇALIŞMALAR

Bu bölümde doğal dil işleme ve duygu analizi konusunda yapılmış çalışmalar incelenmiştir.

Computational Approaches for Emotion Detection in Text

Haji Binali, Chen Wu ve Vidyasagar Potdar tarafından Curtin Teknoloji Üniversitesi bünyesinde yapılan bu çalışmada [1] metinlerde duygu analizi yapılabilmesi hedeflenmiştir. Çalışmanın temelinde insanların ruhsal durumlarına göre sözcükler kullandıkları öngörülmüştür. Örneğin, insanlar mutlularsa pozitif ifadeleri, sinirlilerse negatif ifadeleri kullandıklarına dikkat çekilmiştir. Yapılan çalışmayla da finansal açıdan müşteri ve tüketicilerin davranış modellerinin ve tercihlerinin daha kesin olarak çıkarılabileceği düşünülmüştür [1].

Çalışma, metnin cümlelere ayrıldıktan sonra her bir cümlenin özne, yüklem, nesne, edat, sıfat, bağlaç olarak içerdiği kelimelerin ayrıştırılmasına dayanmaktadır.

Türkçe için yapılmamış bir çalışmadır. Türkçe'ye uyarlanması durumunda baştanbaşa yeni bir cümle yapısıyla çalışılacağı için yöntemin algoritmasının tamamen değiştirilmesi gerekecektir.

Twitter Üzerinde Duygu Analizi

Meriç Meral tarafından Yıldız Teknik Üniversitesi [2] çatısında hazırlanan çalışmada Twitter üzerinden paylaşılan Türkçe tivitlerde ‘Olumlu, Olumsuz, Nötr’ olmak üzere 3 sınıfa ayrılarak duygu analizi yapılabilmesi amaçlanmıştır.

Çalışma kapsamında Twitter API’si kullanılarak 9 farklı alandan Türkçe tivit’ler toplanarak bu 9 alanda etiketlenerek NoSQL veri tabanında saklanmıştır. Toplanan tivitlerden 8321 tanesi gönüllü insanlar tarafından ‘Olumlu, Olumsuz, Nötr’ duygulardan hangilerini taşıdıklarına yönelik etiketlenmişlerdir. Tivitlerin işlenmesi için kelime tabanlı ve n-gram olmak üzere 2 farklı yöntem de kullanılmıştır. Yani ilk olarak kelimeler eksiksiz ve tekrarsız olarak toplanıp sınıflanmıştır. Daha sonraki aşamada da kelimeler ikili ve üçlü kelime grupları olarak gruplandırılmıştır. Son olarak da ‘Rastgele Orman’, ‘Naive Bayes’ ve ‘Destek Vektör Makinesi’ kullanılarak tivitler işlenmiştir. Kelime tabanlı yöntem ile N-gram yönteminden elde edilen sonuçlar ayrı ayrı incelenmiştir.

Çalışmada duygu kategorileri sığ bir alanda incelenmiştir. Direkt duygulara ayrılması yerine negatif ve pozitif duygular gruplandırılmıştır.

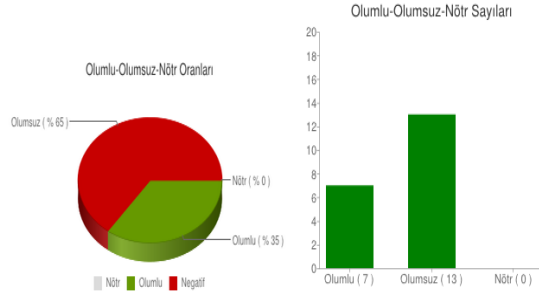
BigDugong – Botego

Bir Türk firması olan Botego tarafından yapılan ‘Sentiment Analizi’ çalışması kapsamında ‘BigDugong’ adını verdikleri bir uygulama oluşturmuşlardır. Uygulama internet üzerindeki sosyal ağlarda paylaşılan Türkçe içeriklerin ifade ettiği duyguları ‘Olumlu-Olumsuz-Nötr’ olmak üzere 3 sınıf altında toplamakta. Uygulamaya ilk olarak belli bir veri kaynağından gelen ifadeler elle öğretilmektedir. Öğretme işlemi yapıldıktan sonra, uygulama artık kendi kendine öğrenme sürecine girer ve işletilmeye hazır olur. [3]

Uygulama Botego tarafından verilen bilgiye göre her hangi bir alan üzerinde yapılan sonuçlarda %70 başarı göstermekte. Öğrenme işleminin gerçekleştirilmesiyle de bu oran %80 oranını geçmektedir. Uygulama temelinde N-gram yöntemini kullanmaktadır.

Çalışmada duygu kategorileri kullanılması yerine duygular direkt pozitif ve negatif gruplar altında toplanmıştır.

"akbank" için sentiment analizi



İçinde "akbank" geçen tweetler

RT @Akbank: Temmuz'da Akbank Direkt İnternet'e gir, unutulmaz bir tatil kazanma şansı yakalal! <http://t.co/xuEp9M3A>

Temmuz'da Akbank Direkt İnternet'e gir, unutulmaz bir tatil kazanma şansı yakalal! <http://t.co/xuEp9M3A>

RT @AkbankSanat: Akbank Sanat Caz Günleri'nin Ardından Micheal Wollny ile Müzik Üzerine <http://t.co/U9tdEKF9>

Akbank Sanat Caz Günleri'nin Ardından Micheal Wollny ile Müzik Üzerine <http://t.co/U9tdEKF9>

birikimlerinizi kesinlikle akbank,a yatırmayın hesabınızdan ne zaman ne kadar kesinti yapacağı belli olmaz. soru sormaya bile para kesiyorlar

akbank kredi sorma ya hesabınızdan habersiz para alıyor. akbank ile hiç bir şekilde çalışmayın...

@meltemgezici evet akbank, deniz, ziraat var.

Doğru Yanlış

#AKBANK madem atm nizin para yatırma ünitesi bozuk niye parayı yatırırken uyarı vermiyorsunuz..

Siz siz olun #AKBANK in atm lerini kullanmayın..

Şekil.1. Botege tarafından yayımlanan örnek duygu analizi verileri

Kaynak: (21.01.2017) tarihinde <http://www.sentimentanalizi.com/sentiment-analizi-uygulamasinin-ekran-goruntusu/> adresinden alındı.

Otomatik Duygu Analizi Sözlüğü Çevirimi ve Duygu Analizinde Kullanımı

Alaettin Uçan tarafından Hacettepe Üniversitesi çatısında yapılan Çalışma kapsamında İngilizce duygu sözlüğünün Türkçe'ye çevirimi yapılmıştır [4].

Çevirimi yapılırken kelimelerin cümlenin hangi ögesi oldukları, pozitif/negatif/nötr duygu bildirme puanları, kelimenin kullanıldığı örnek cümlelerle birlikte tutulmuştur.

Metinlerde geçen her bir kelimenin Türkçe çevirisi için öncelikli olarak 'Google Translate' API'si kullanılmıştır. Bu API ile beraber 3 farklı çeviri sitesinin daha API'leri kullanılmıştır. Çevirilerin API'ler yardımıyla yapılmasında 'Seri çeviri algoritması', 'Paralel çeviri algoritması' ve 'Anlamdaş ilişki' yöntemleri olmak üzere 3 farklı yöntem kullanılmıştır.

'Seri çeviri algoritması' yönteminde, kelimenin Türkçe karşılığı ilk olarak 'Google Translate'te aranmış, burada kelimenin Türkçe karşılığı bulunursa çeviri veri tabanında tutulmuştur. Eğer çevirisi 'Google Translate'te bulunamamışsa da sırasıyla çevirisi bulunana kadar diğer 3 sözlüğün API'leri kullanılmıştır.

'Paralel çeviri algoritması' yönteminde, sözlükler aynı anda birbirlerine paralel olarak çalıştırılmışlardır. Elde edilen sonuçlarda da her hangi en az iki sözlük çevirisi birbirinin aynısıysa çevirinin doğru olarak yapıldığı kabul edilmiş ve veri tabanında bu çeviri tutulmuştur.

'Anlamdaş ilişki' yöntemindeyse, hem 'Seri çeviri algoritması' hem de 'Paralel çeviri algoritması' yöntemleriyle çevirisi bulunamayan kelimelerin 'SentiWordNet' içerisindeki anlamdaş ilişkisinden faydalanılarak karşılıkları bulunmaya çalışılmıştır. Bu doğrultuda kelimenin anlamdaşının varsa 'Seri çevirme algoritması' ile bulunan çevirisi, yoksa da 'Paralel çevirme yöntemi' ile bulunan çevirisi veri tabanında tutulmuştur.

Duygu analizi yapılırken de 'Otel' ve 'Film' yorumları kullanılmıştır. Her bir yorumun öncelikle Zemberek kütüphanesi aracılığıyla yazım hataları düzeltilmiştir.

Ayrıca, yine Zemberek kütüphanesi kullanılarak kelimeler eklerinden arındırılıp sadece kökleri tutulmuştur. Tutulan her bir kökün yorum içerisinde kaç defa geçtiği tutularak frekans sıklıkları bulunmuştur. En yüksek frekans sıklığına sahip olan kökün/köklerin ifade ettiği baskın duygunun o yorumun ifade ettiği duygu olduğu kabul edilmiştir.

Emotion Recognition for Affective User Interfaces using Natural Language Dialogs

Cheongjae Lee and Gary Geunbae Lee tarafından Pohang Bilim ve Teknoloji Üniversitesi kapsamında hazırlanan çalışmada duygular ‘Nötr, Mutlu, Şaşkın, Kızgın, Üzgün, Alaycı, Korkmuş, Sıkkin’ olmak üzere 8 sınıfta toplanmıştır. İçeriklerin dilsel açıdan işlenmesinde ikili n-gram yöntemi kullanılmıştır [5].

Hazırlanan sistemin öğrenme aşamasında da içeriklerin ifade ettiği duygular için kullanıcı geri dönüşleri kullanılmıştır.

Metin içerisinde ağır basan duygunun bulunmasında da metin içerisinde kelimenin kullanım sıklığı ve aynı kategorideki diğer metinlerde kelimenin kullanım sıklığına bakılmıştır. Ayrıca kelimenin kullanım yeri dikkate alınarak kelimenin göze çarpıcılığı da göz bulundurulmuştur.

Identifying Expression of Emotions in Czech Text Using Semantic Relations for Dimension Reduction

Redek Cervenc ve Radim Burget tarafından Brno Teknoloji Üniversitesi bünyesinde hazırlanan çalışmada [6] ilk aşamada manuel olarak her bir örnek metnin hangi duyguyu ifade ettiğini barındıran bir veri seti oluşturulmuştur. Makine öğrenmesi aşamasında bu hazırlanan veri seti kontrol bloğu olarak kullanılmıştır. Duygular, 'Pozitif, Nötr, NegatifL1, NegatifL2, NegatifL3, Kaba' olarak sınıflandırılmıştır. Her bir sınıfın da ifade ettiği puanlamaya göre kontrol grubu içerisinde kelimelerin kullanım frekansına bakılarak deney grubunun da ifade ettikleri duygular tespit edilmeye çalışılmıştır.

OMG! How do you know my feeling? A predictive model to determine sentiments from Tweets

Yi Zhong ve Zefeng Guo tarafından Kaliforniya Üniversitesi bünyesinde hazırlanan çalışmada Twitter sosyal medya ağında paylaşılan tivitlerde duygu analizi yapılabilmesi amaçlanmıştır[7].

Çalışma kapsamında duygular 'Pozitif, Negatif, Nötr' olmak üzere 3 sınıfa toplanmıştır. Twitter üzerinden veriler çekilirken de hashtag'li olanlar yani tivitın önünde # (diyez) işaretiyle birlikte bir ifade barındıran tivitler toplanmıştır. Ayrıca duyguların saptanabilmesi için emoji sözlüğü de oluşturulmuştur. Deney grubunda kullanılan her bir tivitın duygu analizi için de 'n-gram' ve 'olasılık' metotları kullanılmıştır.

3. KAVRAMLAR

Bu bölümde çalışmada kullanılan teknoloji ve kavramlar açıklanmıştır.

3.1 Duygu Analizi

Duygu Analizi ya da Sentiment Analiz temel olarak yazılı ya da sözlü bir içerikte belirtmek istenen duygunun analiz edilmesidir. Duygu analizi özellikle firmaların ürünleri hakkında kullanıcıların fikirlerini almak için internet üzerinde geniş bir alanda kullanılmaktadır.

3.2 Twitter

Twitter, nam-ı diğer ‘İnternetin SMS’i, bütün dünya üzerinde yaygın olarak kullanılan bir sosyal paylaşım sitesidir.

Diğer sosyal paylaşım sitelerinden farklı olarak kullanıcılar Twitter üzerinde her biri en fazla 140 karakterlik ‘Tivit’ adı verilen mesaj blokları paylaşırlar. Tivitlerin yayınlanma sıklığı ya da yayınlanma zamanı gibi her hangi bir kısıt olmadan kullanıcılar istedikleri kadar tivit yayımlayabilirler. Atılan tivitler bütün Twitter kullanıcılarına açık olmakla beraber kullanıcılar tivitlerini hangi kullanıcıların görmek istediği şekilde sınırlayarak hesaplarını konfigüre edebilirler. Ayrıca, her bir tivitinde haberdar olmak istediği kullanıcıların ya da kurumların tivitlerine üye olabilir ve bu kişi / kurumların takip ettiği kişi/kurumlar ile başka hangi kullanıcılar daha tarafından takip edildiklerini görebilirler.

Kullanıcılar twitter'a yalnızca web sitesi üzerinden değil, aynı zamanda web sitesi ile uyumlu harici uygulamalar ve bazı kısa mesaj servisleri tarafından da erişebilirler.

Günümüzde dakika başına ortalama doksan sekiz bin adet tivit atılan Twitter sosyal ağı 2006 senesinde Amerika tabanlı bir firma olan 'Odea' tarafından hayata geçirilmiştir. Jack Dorsey'nin öncülüğünde Biz Stone, Evan Williams ve Noah Glass tarafından Odea bünyesinde insanların durumlarını mesaj olarak paylaşabileceği bir proje üzerinde çalışmaya başlamalarının sonucu ortaya çıkan bir projedir Twitter. [8]

2011 senesinden itibaren de Twitter artık Türkçe desteği de sunmaya başlamıştır.

Twitter günümüzde artık hayatın o kadar olması gereken doğal bir parçası haline geldi ki kendisine özgü kısaltmaları da bütün dünyada insanlar tarafından benimsenmiş durumdadırlar. Twitter'ın kendine özgü başlıca 4 tane kısaltması bulunmaktadır.

Bunlardan ilki 'TT' yani Trending Topic'in kısaltmasıdır. Trending topic'ler adından da anlaşılacağı üzere gündemi belirleyen, insanların üzerine en çok paylaşımda buldukları başlıklardır. Twitter'ın ana sayfasına girildiğinde sol tarafında listelenen 10 adet konudur. Trending topic'ler ülkelere özel olarak değiştiği gibi konular aracılığıyla bulunulan şehir/bölgelere özel olarak da gösterilmektedir.

Bir diğer twitter kısaltması ise 'RT' yani Retweet'dir. Retweet de başka bir kullanıcı tarafından alıntılanmış olan tivitlere verilen özelleştirilmiş bir addır. Kullanıcılar retweet yapmak istedikleri başka kullanıcıların tivitlerini, tivitinin yanında bulunan 'Retweet' butonuna basarak yapabilirler. Retweet yapılan tivitler, retweet yapan kullanıcının sayfasında 'KullanıcıAdı+Retweetledi' başlığıyla ve tivitinin asıl sahibinin kullanıcı adı ile sunulur.

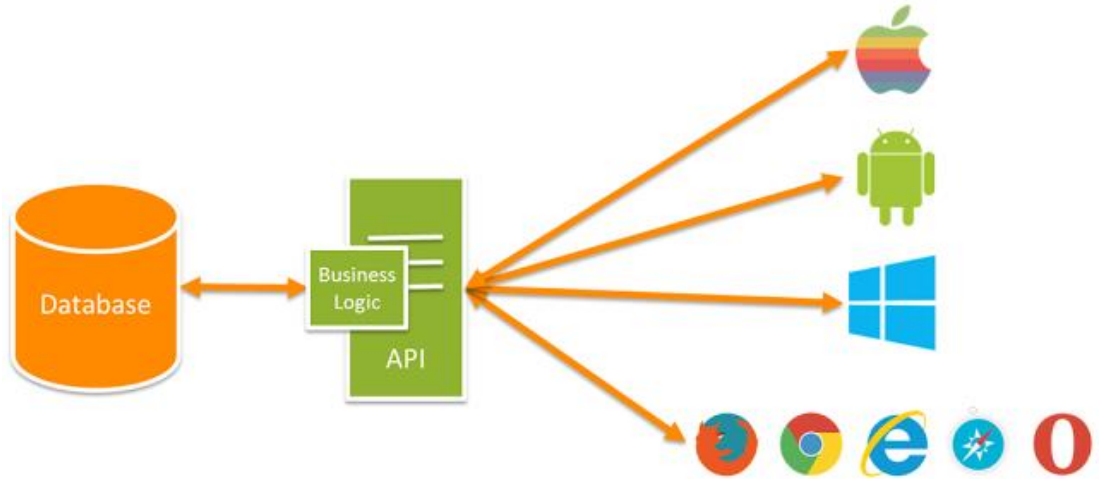
Üçüncü bir kısaltma olan 'FF' ise Friday follow'un kısaltmasıdır. Her ne kadar bu kısaltma 'Cuma takibi' anlamına da gelse haftanın her günü yapılabilir. Friday follow kullanıcıların hoşuna giden ve takipçilerinin de hoşuna gideceklerini düşünüp onlarla da paylaşmak istedikleri kullanıcı/kuruluşların hesabını takibi

listelerine duyurmak için kullandıkları bir kısaltmadır. Takipçilerine bu şekilde paylaşmak istedikleri kişi/kuruluş hesaplarını kullanıcılar '#ff' hashtag'yle paylaşırlar.

Sonuncu twitter kısaltması ise 'DM', yani Direct Message'dır. Direct message da yine adından anlaşılacağı üzere twitter kullanıcıları arasında birbirlerine özelden gönderdikleri mesajlara verilen addır.

3.3 Uygulama programlama ara yüzü (API)

Uygulama programlama ara yüzü (Application programming interfaces - API), yazılımlar arası veya veri tabanları arası iletişimi sağlayan yapılardır. Örneğin, Twitter API'si kullanılarak Twitter üzerinde tivit paylaşımı ya da paylaşılmış olan tivitlerin çekilmesi gibi işlemleri Twitter kullanmadan hazırlanan bir yazılım üzerinden gerçekleştirilir.



Şekil.2. API Mantığı

Kaynak: (21.01.2017) tarihinde <https://lbozan.blogspot.com.tr/2016/09/aspnet-web-api.html> adresinden alındı.

3.4 Twitter API

Twitter API'si twitter verisine programatik olarak okuma ve yazma yapılmasını sağlar. Kısacası, Twitter'ı Rest API'lerine ve Streaming API'lerine ulaşabilmek için kullanılan Phyton paketidir.

Twitter'ın API'sini kullanarak yazılım geliştiriciler Twitter'ın çekirdek verilerine ulaşım sağlayabilirler. Örneğin kullanıcı bilgileri ve kullanıcının takipçi bilgilerinin alınması sayılabilir. Aşağıda Türkiye üzerinden paylaşılan tivitlerin çekilmesi için hazırlanmış bir kod parçası bulunmaktadır [9].

```
1. r = api.request('statuses/filter', {  
2.   'locations': '25,44,35,42'  
3. }) for item in r: print(item)
```

Ayrıca, Twitter üzerinde çoklu paylaşımında bulunma gibi kullanıcı işlemleri de API'ler aracılığıyla otomatik olarak gerçekleştirilebilir.

Örneğin, aşağıda verilen kod parçasıyla Twitter üzerinde 'Merhaba' içerikli bir paylaşım yapılır.

```
1. r = api.request('statuses/update', {  
2.     'status': 'Merhaba '})  
3.     print(r.status_code)
```

Diğer sosyal medya sitelerinde de olduğu gibi, Twitter API'si kullanılarak haberleşme sağlayabilmek için yetkilendirme amacıyla bir key alınması gerekir. Twitter API'si yetkilendirmede OAuth kullanmaktadır.

3.5 OAuth

OAuth, kullanıcıların şifre paylaşımlarına gerek olmadan, web, mobil ve masaüstü uygulamalardan kullanılabilen bir yetkilendirme yani kimlik doğrulama protokolüdür. Açık bir protokoldür.

2010 yılından itibaren bütün Twitter uygulamaları için zorunlu olan OAuth protokolünün ilk ihtiyaç sinyalleri 2006 senesinde Blaine Cook tarafından Twitter OpenID uygulamasını geliştirirken doğdu. İhtiyacın doğmasının ardından 2007 senesinde küçük bir OAuth tartışma grubu kuruldu. 2010 senesinde de OAuth 1.0 protokolü yayınlandı [10].

3.6 JavaScript Nesne Notasyonu (JSON)

JavaScript Nesne Notasyonu (JSON) Java script tabanlı bir veri taşıma formatıdır. XML'in alternatifi olan JSON'ın XML'e karşı avantajı XML'den daha az yer kaplamasıdır. Ayrıca, JSON insanlar tarafından kolaylıkla okunulabilir ve insanlar tarafından kolaylıkla yazılabilir bir formattır. Aynı şekilde JSON bütün programlama dilleri ile yazılmış olan yazılımlar tarafından kolaylıkla anlaşılabilir ve bu yazılımlar tarafından kolaylıkla üretilebilir bir formattır [11].

JSON key yani anahtar ve value yani değer çiftleri formatında ifade edilir. Örneğin, aşağıda verilen örnek JSON formatında soldaki ifadeler (Ad, Soyad, Bölüm, Üniversite ve Numara) anahtarlar, sağdaki ifadeler de (Burcin, Adak Kaplan, Bilgisayar Muhendisligi, Beykent Universitesi ve 130820021) değerlerdir.

```
1.  {  
2.    "Ad": "Burcin",  
3.    "Soyad": "Adak Kaplan",  
4.    "Bolum": "Bilgisayar Muhendisligi",  
5.    "Universite": "Beykent Universitesi",  
6.    "Numara": "130820021"  
7.  }
```

3.7 Linq2Twitter

Linq2Twitter, Twitter API'sinin C# programlama dili için geliştirilmiştir. Açık kaynak kodlu bir yazılımdır [12].

Linq2Twitter Örneği

```
1. Console.WriteLine("Tivitleri al: ");
2. try {
3.     statusTweets = (from tweet in twitterContext
                       where tweet.Type == SearchType.Search && tweet.Query == searchTweet
                       && tweet.Count == tweetCount
                       select tweet).ToList();
4. }
5. catch (AggregateException e) {
6.     Console.WriteLine("Hata {0}", e.InnerExceptions[0]);
7. }
8. private static void PrintTweets(List < Search > statusTweets)
{
9.     foreach (var statusTweet in statusTweets) {
10.        statusTweet.Statuses.ForEach(tweet =>
11.            Console.WriteLine("KULLANICI: {0}, Tweet: {1} \n",
tweet.User.ScreenNameResponse, tweet.Text));
12.        Thread.Sleep(1000);
13.    }
14. }
```

3.8 NoSQL

NoSQL (Not Only SQL), yani SQL ve daha fazlası klasik ilişkisel veri tabanı yönetim sistemlerine alternatif olarak geliştirilmiştir. [13]

1998 senesinde çıkartılan NoSQL, Carlo Strozzi tarafından tasarlanmıştır. [14]

BaseX, CouchDB, eXist, Jackrabbit, Lotus Notes, MarkLogic Server, MongoDB, SimpleDB ve Terrastore doküman tabanlı NoSQL uygulamalarına örnektir. [14]

AllegroGraph, Neo4j, Dex ve FloxDB de Çizelge tabanlı NoSQL uygulamalarına örnektir.

3.9 DocumentDB

DocumentDB Microsoft firmasının bulut ürünü olan Azure üzerinden İnternet ortamında bulunan JSON formatlı belgelerin yönetilebilmesi için sunduğu çok kullanıcıya bir NoSQL çözümdür.

İlişkisel veri tabanı yoktur. .Net, .Net core, Java, Python ve node.JS ortamları tarafından desteklenmektedir. Ayrıca Rest Api ve MongoDB aracılığıyla da erişilebilir [15].

3.10 Zemberek Kütüphanesi

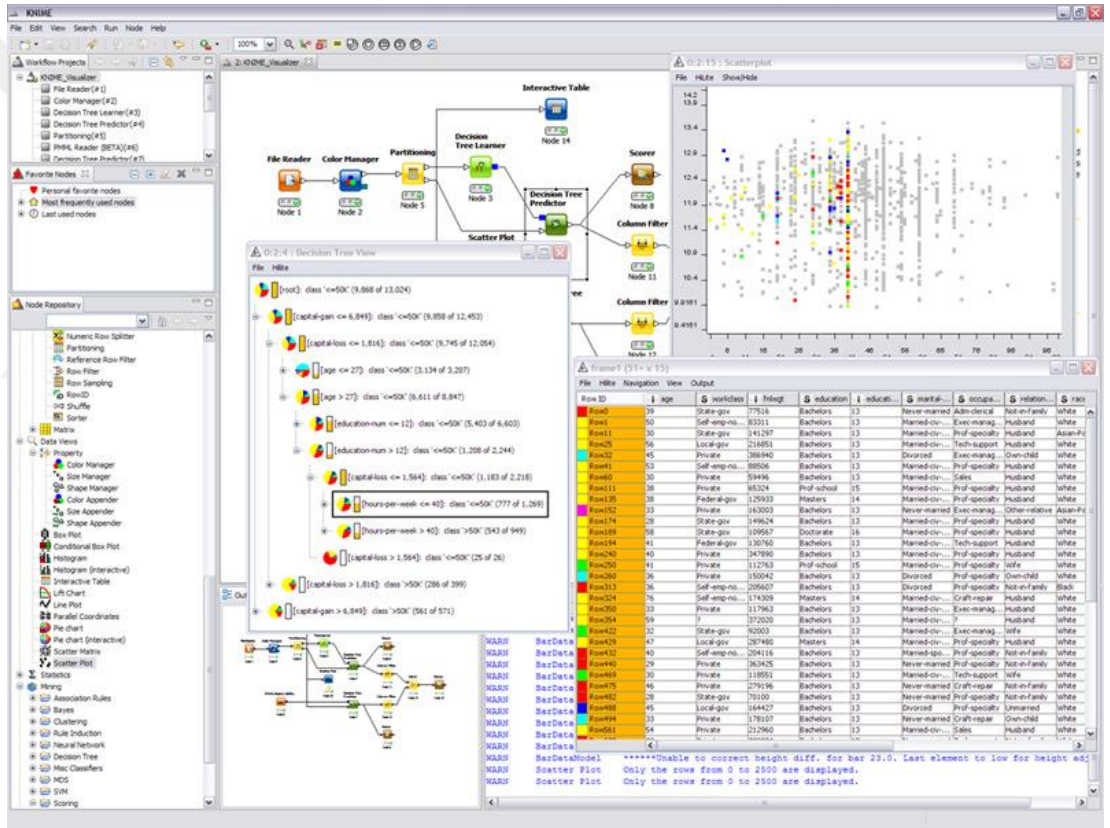
Zemberek, açık kaynak kodlu ve Java ile geliştirilmiş bir kütüphanedir. Zemberek kütüphanesi kullanılarak Türkçe 'deki yazım hataları düzeltilebileceği gibi, kelimeler ek ve köklere ayrılabilirler. Yani, Zemberek bir Türkçe Doğal dil işleme kütüphanesidir. Ayrıca OpenOffice ve LibreOffice eklentisidir [16].

2005 senesinde en iyi Özgür Yazılımı ödülünü almıştır.

3.11 Knime

Knime (Konstanz Information Miner) açık kaynak kodlu bir veri analizi, raporlama ve entegrasyon platformudur.

2006 senesinde ilk versiyonu sunulan Knime'nin geliştirilmesine 2004 senesinde Konstanz Üniversitesi'nin yazılım mühendisleri ekibi tarafından başlanmıştır [17].

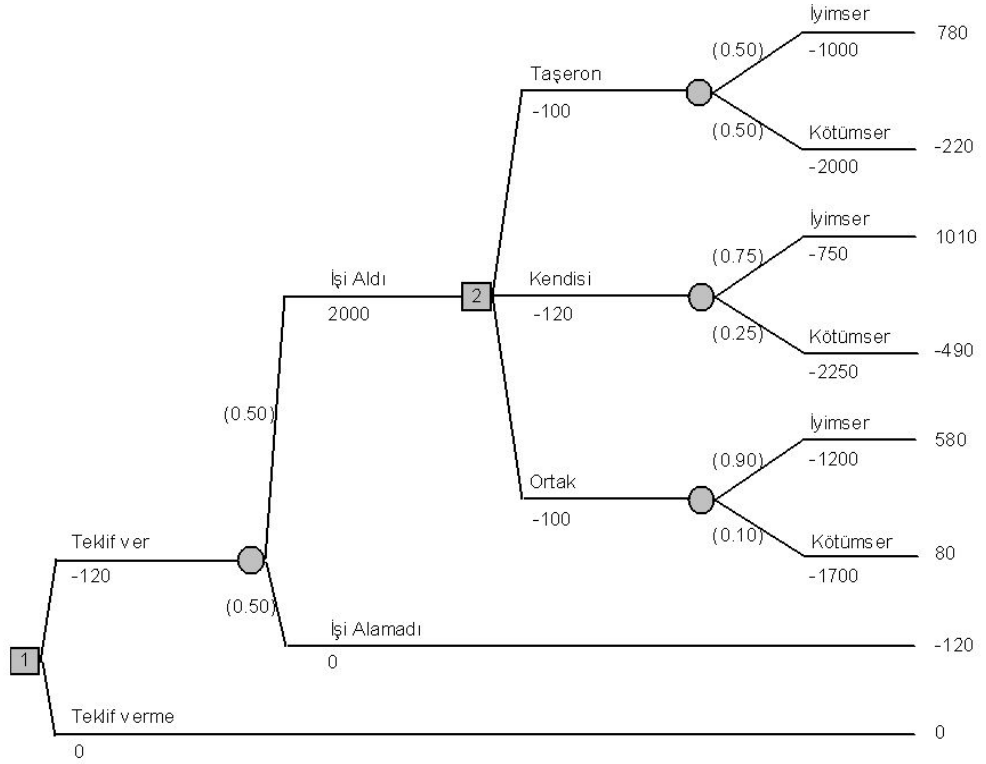


Şekil.3. Knime Örnek Ekran Görüntüsü

Kaynak: (22.01.2017) tarihinde <https://en.wikipedia.org/wiki/KNIME> adresinden alındı.

3.12 Karar Ağacı

Öğrenme elemanlarından karar ağaçları yöntemi en güçlü ve en yaygın sınıflandırma ve öngörü araçlarından birisidir. Ağaç yapılı yöntemlerin sık kullanılmasının nedeni ise yapay sinir ağlarının tersine ağaç yapılarının kuralları ifade edebilmesinden kaynaklanmaktadır. Oluşan kurallar çok sade ve nettir. Eğer Ağaç çok fazla büyük değilse her insanın kolaylıkla okuyup anlayacağı türden bir sınıflandırma gerçekleştirir.



Şekil.4. Karar Ağacı Örneği

Kaynak: Karar Ağaçları, Doç. Dr. Aydın Ulucan (20.01.2017) tarihinde

<http://slideplayer.biz.tr/slide/2311852/> adresinden alındı.

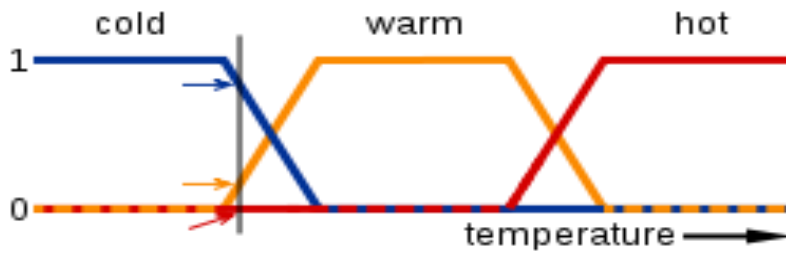
3.13 Bulanık Mantık

Bulanık mantık (Bulanık eseme, Puslu mantık) ilk olarak 1965 senesinde bilimadamı Lotfi Aliasker Zadeh tarafından önerilmiştir.

Bulanık mantık küme ve altküme yapılarına göre şekillendirilir. Bulanık mantığa göre varlıklar üyelik derecesine göre değerlendirilir. Üyelik dereceleri 0 ile 1 arasında her hangi bir değer olabilir. Yani klasik kümelerden farklı olarak bulanık mantık kümelerinde üyeler sonsuz değer alabilir [18].

Ayrıca klasik mantık ile bulanık mantık arasındaki farkları ele alırsak;

- Klasik mantıkta değer kesin, bulanık mantıkta değer kısmidir.
- Klasik mantık 'A veya A değil' ifadesi Bulanık mantıkta 'A ve A değil' halini alır.
- Klasik mantıkta ya hepsi ya hiçbiri iken (ya var ya yok) bulanık mantıkta belirli derecelerde varlık kabul edilir.
- Klasik mantıkta varlıklar ikili birimlerle ifade edilirken, bulanık mantıkta varlıklar bulanık birimlerle ifade edilir.



Şekil.5. Bulanık Mantık

Kaynak: (20.01.2017) tarihinde

https://tr.wikipedia.org/wiki/Bulan%C4%B1k_mant%C4%B1k adresinden alındı.

3.14 C# programlama dili

Microsoft firması tarafından geliştirilmiş olan C#, .Net uygulamalarını geliştirmek için 2000 senesinde çıkartılmış olan bir programlama dilidir. Üçüncü nesil, üst seviye bir programlama dilidir. Nesne yönelimli yani object oriented bir dildir.

C++ ve Java dillerinden etkilenmiştir ve kendisi de D, Dart, Swift, Java dillerini etkilemiştir.

Örnek c# programı (Basit hesap makinesi)

```
using System;

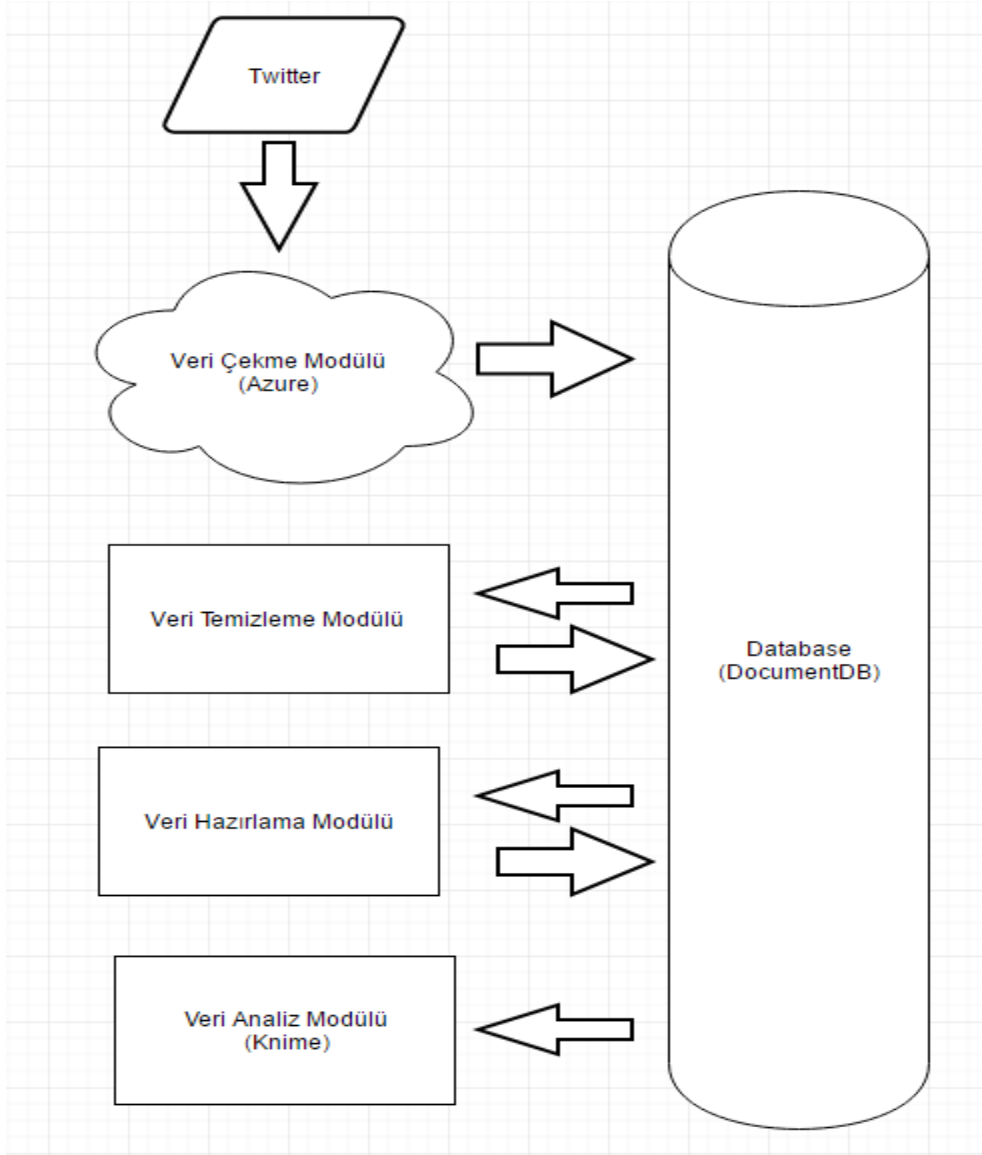
class Program
{
    1.  {
    2.  int sayi1, sayi2, sonuc;
    3.  string islem;
    4.  static void Main(string[] args) {
    5.  Console.WriteLine("İlk sayı: ");
    6.  sayi1 = Convert.ToInt32(Console.ReadLine());
    7.  Console.WriteLine("İslem: ");
    8.  islem = Console.ReadLine();
    9.  Console.WriteLine("İkinci sayı: ");
    10. sayi2 = Convert.ToInt32(Console.ReadLine());
    11.
    12. if (islem=="-")
    13. { sonuc = sayi1 - sayi2;}
    14.
    15. else if (islem=="+")
    16. {sonuc = sayi1 + sayi2;}
    17.
    18. else if (islem=="/")
    19. {sonuc = sayi1 / sayi2;}
    20.
    21. else if (islem=="*")
    22. {sonuc = sayi1 * sayi2;}
    23.
    24. else
    25. {cevap = 0;}
    26.
    27. }
    28. Console.WriteLine(sayi1.ToString() + " " + islem + " " + sayi2.ToString() +
    " = " + sonuc.ToString());
    29. Console.ReadLine();
    30. }
    31. }
```

3.15 Veri Madenciliği

Veri madenciliği, büyük ölçekli verilerin işlenerek bilgiye ulaşılması işlemidir. Bu sayede bilgisayarlar aracılığıyla işlenen büyük miktarlarda veri sayesinde bir sonuca etki eden nedenler kolaylıkla ayrıştırılabilir ve analizi yapılır. Ayrıca geleceğe yönelik durum olay tahminlemeleri yapılabilir [19].

4. YÖNTEM

Bu bölümde Türkçe için duygu analizi programının oluşturulması ele alınmıştır. Bölümün ilk kısmında kullanılan kavramlar açıklanmış, ikinci kısmında ise yöntem aşamaları anlatılmıştır.



Şekil.6. Duygu Analizi Programı Aşamaları

4.1 Verilerin Toplanması

Çalışma kapsamında ilk olarak Twitter’da Türkçe olarak paylaşılan tivitlerin çekilmesi hedeflenmiştir. Bunun için Linq2Twitter yazılımını ve Twitter Streaming API’sini kullanan bir program geliştirilmiştir. Program Twitter Streaming API’sine kayıt olarak Türkçe tivit geldikçe, bu tivitleri dinleyerek kaydedecek şekilde programlanmıştır.

Programın uzun süreli ve durmadan çalışması gerekmektedir. Elektrik kesintisi, güç kaybı yaşanması ve programlama hataları gibi durumlarda veri kaybı yaşamamak için program bulut platformuna taşınmıştır. Geliştirilen programın kesintisiz çalışması için Microsoft Azure platformu üzerinde bir servis görevi oluşturulmuştur. Böylelikle geliştirilen yazılımın sürekli olarak çalışması garanti edilmiştir. Program 01.05.2015-15.06.2015 tarihleri arasında çalışarak veri toplama işlemini yapmıştır. Toplamda bu verilen çalışma süresi içerisinde 989583 adet Türkçe dilinde paylaşılmış tekil tivit çekilmiştir.

Toplanan verinin çokluğu ve hızla işlenmesi gerektiği için klasik ilişkisel veri tabanı kullanılmasının yeterli olmayacağı ön görülmüştür. Bu nedenle program sonucu elde edilen veriler bir NoSQL veri tabanı olan DocumentDB üzerine kaydedilmiştir.

```
TwitterTestConsoleApplication
Program.cs
TwitterTestConsoleApplication
TwitterTestConsoleApplication.Program
SaveTweets(TwitterContext twitterCtx)

1 using System;
2 using System.Collections.Generic;
3 using System.IO;
4 using System.Linq;
5 using System.Text;
6 using System.Threading.Tasks;
7 using LinqToTwitter;
8 using Newtonsoft.Json;
9 using TwitterTestConsoleApplication.Models;
10
11 namespace TwitterTestConsoleApplication
12 {
13     class Program
14     {
15         private static string _twitterApiKey = "uM3cX0V3X0W6U8H9wefC0M";
16         private static string _twitterApiSecret = "U8v9Yh11hAggCQ0MacCCxIu8wMHTpSuzi1h17e1s";
17         private static string _twitterAccessToken = "2956141436-qf94P4hT07279fnsXQ0K4Cz0u6uG9K1rv0L";
18         private static string _twitterAccessSecret = "T1kxvryQ0MEFa2I24Jz0mP40EN056JL0U33XZJh9mC";
19
20         static void Main(string[] args)
21         {
22             var auth = new SingleUserAuthorizer
23             {
24                 CredentialStore = new SingleUserInMemoryCredentialStore
25                 {
26                     ConsumerKey = _twitterApiKey,
27                     ConsumerSecret = _twitterApiSecret,
28                     AccessToken = _twitterAccessToken,
29                     AccessSecret = _twitterAccessSecret
30                 }
31             };
32             var twitterCtx = new TwitterContext(auth);
33             try
34             {
35                 SaveTweets(twitterCtx).GetAwaiter().GetResult();
36             }
37             catch (Exception)
38             {
39                 twitterCtx.Dispose();
40                 Main(null);
41             }
42         }
43     }
44
45     static async Task SaveTweets(TwitterContext twitterCtx)
46     {
47         int count = 0;
48         int iteration = 1;
49         TweetContext context = null;
50
51         await (from str in twitterCtx.Streaming
52              where str.Type == StreamingType.Sample && str.Language == "tr"
53              select str
54              ).StartAsync(async str =>
55             {
56                 if (count++ % 100 == 0)
57                 {
58                     Console.WriteLine($"-----Iteration {iteration} started-----");
59                     context = new TweetContext();
60                 }
61             });
62     }
63 }
```

Şekil.7. Veri Çekme Modülü Kodlar Bölüm 1

Tablo.1. Örnek Çekilen Tivitler

1	RT @SavciSayan: Hayırlı Cumalar... https://t.co/2BEYCyAdVH
2	Galatasaray'ın yeni hocası belli oldu! https://t.co/inljnyUGk1
3	Tirnagim kırılmış hir cıkaririm bana bulasabileceğim birini rtle dusurun
4	RT @SOZ_KIRINTILARI: ___ GÜNAYDIN ❤️ Günümü aydınlatan ve bu güne beni erdiren Rabbime şükürler olsun. Hayırlı Nur'lu Cuma/lar ...
5	RT @cicek_meliha: ""Ben Gibi"" Olan Herkesten...!!! ""Sen Gibi"" Olan Herkese Gelsin...!!! (Taladro = Sen Gibi)
6	RT @Kralice_Hanim: En çok ilgi gören ve satan kitaplar arasında zirvede yolculuğuna devam ediyor #BudaGeçerYaHu @FerudunOzdemir https://t.co/...
7	RT @Sinekovski: Biz umudumuzu sadece Allah'a bağlarız..
8	RT @MunirUstun: Çok yakında siz okuyucularımızı ve dostlarımızı karşılayacağımız mekanın çizimleri geldi... @profilkitap https://t.co/ajOve...
9	RT @lebdeb: #KalbimSamsunda BEDENİM DEĞİL AMA RUHUM ORADA. BAŞIMIZ SAĞ OLSUN. https://t.co/KIbCd8QKGV
10	RT @karlev5806: LOCA Efendinin Donunu Koklayıp, Sırtını RESULULLAH'a, Yüzünü KANDİL'e Dönen Kripto HİMMETÇİLER !.. @MahirAytekin https://t.co/...
11	RT @av_alperyilmaz: @gokhanyilmaz_hr ve @BurcinSORAYERDG'in son çalışması özeleştiriyapmanın vaktinin geldiğini gösteriyor. Kendi blogumu ...
12	RT @sputnik_TR: #Mali Radisson'da saldırı: 170 kişi rehin alındı https://t.co/M1cXaaJi9C https://t.co/98cpOumGc4
13	@7321_21 @evin_kurdistan ne yazık ki ..
14	@KeremBursin gurur kaynağımızsın be adam 😊😊 https://t.co/euNoBArQ1Z
15	RT @asya_alt: .. Her şey, ama her şey eskiye kaldı Vakit yok bir daha yemyeşil Eylül tramvaylarına. ... #edipcanserver https://t.co/RrrWosp...
16	RT @bedohas: Bugün Rahmet.. Bugün Bereket.. Bugün Şefaat.. Bugün Himmet.. Bugün günlerin Efendisi.. CUMAMIZ MÜBAREK OLSUN... https://t.co/w...
17	Bulvar Yaşam ve Alışveriş Merkezi 2016 Yılbaşı Kampanyası https://t.co/mZEu9XJ5e6 @kimkazandi aracılığıyla
18	RT @yazarsecmeleri: "Dua ettiğinizde çok isteyiniz çünkü vereceği hiçbir şey Allah'a ağır gelmez." /Hz. Muhammed s.a.v/
19	Fatma Nur'un ders çalışması gerekiyordu ama olmadı 😊 (@ Gazi Mustafa Kemal Devlet Hastanesi in Ankara, Türkiye) https://t.co/Pv25NTSHWo
20	Gezici Festival, 21. Yolculuğuna Çıkıyor {haberler} https://t.co/P8P07aMQIm
21	RT @KeskinSozler: "Mutluyken bir anda üzgün moda girmeniz, çok büyük ihtimalle birisini çok fazla özlediğinize işarettir." – Anonim
22	@ruzgarsagnak saolun başkanım
23	BEN SENİN KOLLARINI VATAN BİLMİŞİM
24	@SideequeRajput @macsingl75 @khalidsalmani1 @SarikaPost @ANI_news Bilkul
25	@shakirdisabr bre cahil siz biz değil Rabbim müslümanları Kuran ve sünnetten ayırmasın ne rte si ne Hocaefendi'si açın gözünüzü
26	RT @Sagopakajmerrap: Periscope yayını 11.30 da !... Trafikteyiz:) İstanbul:)
27	Biz sadece frikik kullanırken geri adım atıyoruz kardeş S.A
28	Uf kalyom 😊
29	RT @HergunYeniBilg: Çocuklarla insanın ruhu tedavi olur... İçten gülen bir çocuk insanın ömrünü uzatır.' Dostoyevski https://t.co/k3fhEf...
30	Özgürlükler Beldesi Tehlikede Mi? https://t.co/z95QteSDSX

4.2 Verilerin Temizlenmesi

Twitter’da Tivitlerin 140 karakterle sınırlı olması kullanıcıların istediklerini tam olarak ifade edebilmek için çok sayıda kısaltma kullanmasına ve yazım hataları yapılmasına yol açmaktadır. Ayrıca Twitter’ın yapısı gereği tivitlerde ‘#’ karakteri ile tanımlı etiketler ve @ ile başlayan kişi isimleri bulunmaktadır. Bu nedenle toplanan verilerdeki yazım hatalarını gidermek ve kullanılmayacak alanların temizlenmesi gerekmektedir.

Kullanılmayacak alanları temizlemek ve yazılım hatalarını düzeltmek için Zemberek kütüphanesini kullanan bir yazılım geliştirilmiştir. Bu yazılım kaydedilen tivitleri tek tek okur. Okunan her tivit için aşağıdaki adımları gerçekleştirir.

Tivit boşluk karakteri ile kelime dizisine ayrılır. ‘@’ karakteri kişiyi ‘#’ karakteri etiketi ifade ettiği için bu harfler ile başlayan kelimeler dikkate alınmaz. Diğer kelimeler Zemberek Kütüphanesi kullanılarak Türkçe olup olmadığına bakılır. Eğer kelime Türkçe değil ise yine Zemberek Kütüphanesi kullanılarak hatalı yazım düzeltilir. Bütün kelimeler tamamlandıktan sonra düzeltilmiş olan tivit DocumentDB üzerine yazılır.

```

TwitterTestConsoleApplication
Program.cs
TwitterTestConsoleApplication
TwitterTestConsoleApplication.Program
SaveTweets(TwitterContext twitterCtx)
1 using System;
2 using System.Collections.Generic;
3 using System.IO;
4 using System.Linq;
5 using System.Text;
6 using System.Threading.Tasks;
7 using LinqToTwitter;
8 using Newtonsoft.Json;
9 using TwitterTestConsoleApplication.Model;
10
11 namespace TwitterTestConsoleApplication
12 {
13     [reference]
14     class Program
15     {
16         private static string _twitterApiKey = "kzP3c0LX3WV4BjHMeefC0d";
17         private static string _twitterApiSecret = "UR9YHq11arAggG0PwCCcGwBwMFTpeSollU7e1s";
18         private static string _twitterAccessToken = "295614136-q7M9PwUd7Z78YhXQ@QzCv1m6uXGK11v0L";
19         private static string _twitterAccessTokenSecret = "1IXvVryQ10kFz1X6j2m9PwUd7Z78YhXQ@QzCv1m6uXGK11v0L";
20
21         [reference]
22         static void Main(string[] args)
23         {
24             var auth = new SingleUserAuthorizer
25             {
26                 CredentialStore = new SingleUserInMemoryCredentialStore
27                 {
28                     ConsumerKey = _twitterApiKey,
29                     ConsumerSecret = _twitterApiSecret,
30                     AccessToken = _twitterAccessToken,
31                     AccessTokenSecret = _twitterAccessTokenSecret
32                 }
33             };
34             var twitterCtx = new TwitterContext(auth);
35             try
36             {
37                 SaveTweets(twitterCtx).GetAwaiter().GetResult();
38             }
39             catch (Exception)
40             {
41                 twitterCtx.Dispose();
42                 Main(null);
43             }
44         }
45
46         [reference]
47         static async Task SaveTweets(TwitterContext twitterCtx)
48         {
49             int count = 0;
50             int iteration = 1;
51             TweetContext context = null;
52
53             await (from strm in twitterCtx.Streaming
54                  where strm.Type == StreamingType.Sample && strm.Language == "tr"
55                  select strm
56                  ).StartAsync(async strm =>
57             {
58                 if (count++ % 100 == 0)
59                 {
60                     Console.WriteLine($".....Iteration {iteration++} started.....");
61                     context = new TweetContext();
62                 }
63             });
64         }
65     }
66 }

```

```

TwitterTestConsoleApplication
Program.cs
TwitterTestConsoleApplication
TwitterTestConsoleApplication.Program
SaveTweets(TwitterContext twitterCtx)
61
62
63     try
64     {
65         await Task.Run(() =>
66         {
67             Status tweet = (Status)strm.Entity;
68             Tweet t = new Tweet()
69             {
70                 Text = tweet.Text
71             };
72             context.Tweet.Add(t);
73             Console.WriteLine(tweet.Text + "\n");
74         });
75
76         if (count++ >= 1000000)
77         {
78             strm.CloseStream();
79         }
80     }
81     catch (Exception)
82     {
83     }
84
85
86
87     if (count % 100 == 0)
88     {
89         await context.SaveChangesAsync();
90         context.Dispose();
91         Console.WriteLine($".....Iteration {iteration-1} ended.....");
92         Console.WriteLine("");
93         Console.WriteLine("");
94         Console.WriteLine("");
95         Console.WriteLine("");
96     }
97     });
98 }
99
100 }
101
102

```

Şekil.9. Veri Temizleme Modülü Kodlar

Verilerin temizlenmesi modülü için hazırlanan yazılımın aşağıda verilen ve ayrıca Şekil 9’da görülen PreRender() metoduyla ilk olarak bütün tivit içerikleri küçük harfe dönüştürülüyor. Türkçe’de büyük harfle ya da küçük harfle yazılması durumunda içerikte bir anlam değişikliği olmamasına rağmen bütün içeriklerin analiz için sokulacağı Knime platformunda analiz esnasında büyük harflerle yazılan bir kelimenin anlamı küçük harflerle yazılan aynı kelimedenden farklı bir kelimeymiş gibi algılanır. Bu nedenle içeriklerin hepsinin aynı formatta olması için hepsi küçük harfe dönüştürülmektedir.

PreRender() Metodu

```
1. public string PreRenderText(string text) {
2.     text = text.ToLower(_culture);
3.     if (text.StartsWith("rt")) {
4.         text = text.Substring(2);
5.     }
6.     var kelimeler = text.Split(' ');
7.     StringBuilder s = new StringBuilder();
8.     foreach(var kelime in kelimeler) {
9.         if (!kelime.StartsWith("@") && !kelime.StartsWith("http://") && !kelime.Sta
rtswith("https://")) {
10.            s.AppendFormat("{0} ", kelime);
11.        }
12.    }
13.    return s.ToString().Trim();
14. }
```

PreRender metodu ile içeriklerin küçük harfe dönüştürülmesinden sonra ilk olarak içeriğin ‘rt’ ifadesiyle başlayıp başlamadığına bakılır. Eğer ifade ‘rt’ ile başlıyorsa bu sadece tivitın bir ‘Retweet’ olduğunu, yani başka bir kullanıcının tivit’inin paylaşılmasıyla oluşturulmuş bir tivit olduğunu gösterir. Dolayısıyla ‘rt’ ifadesi her hangi bir duygu belirtmez. Bu nedenle bu şekilde yakalanan ‘rt’ ifadeleri tivitın içeriğinden silinir.

‘rt’ ifadeleri silindikten sonra da tivitın içeriğinde ‘@’, ‘http://’ ve ‘https://’ ile başlayan kelimeler tivitın içeriğinden çıkartılıyor. Çünkü tivitlerin içeriğinde ‘@’ ile başlayan kelimeler başka bir kullanıcıya etiketleme yapmak amacıyla kullanılır, yani her hangi bir duygu belirtmeyip sadece kullanıcı/hesap adını belirtir. Aynı şekilde ‘http://’ ve ‘https://’ ile başlayan ifadeler de bir web sayfasının adresidir ve her hangi bir duygu belirtmediğinden dolayı bu ifadeler de çıkartılmıştır.

Render() metodu

```
1. public string Render(string text) {
2.     var se = new StringBuilder();
3.     foreach(var t in text) {
4.         if (_alphabet.Contains(t)) {
5.             se.Append(t);
6.         } else {
7.             se.Append(" ");
8.         }
9.     }
10.    return se.ToString();
11. }
```

Yukarıda verilen verilerin temizlenmesi modülü için hazırlanan yazılımın Render() metoduyla tivitin içeriğinde yazılımda tanımlanmış olan ‘alphabet’ dizisinde verilen karakterler haricinde bir karakterin geçip geçmediği kontrol ediliyor. ‘alphabet’ dizisi, ‘A-Z’ arası Türkçe alfabede bulunan bütün harfleri, 0-9 arası rakamları ve ‘@’ işaretini içermektedir. ‘alphabet’ dizisi içerisinde tanımlanmamış bir karakterle karşılaşılması durumunda Render() metoduyla bu karşılaşılan harici karakter tivitin içeriğinden çıkartılır.

```
1. public string Fix(string text) {
2.     var kelimeler = text.Split(' ');
3.     for (int i = 0; i < kelimeler.Length; i++) {
4.         if (!IsTurkish(kelimeler[i])) {
5.             var oneriler = _zemberek.oner(kelimeler[i]);
6.             if (oneriler != null && oneriler.Any()) {
7.                 kelimeler[i] = oneriler.First();
8.             }
9.         }
10.    }
11.    return string.Join(" ", kelimeler.Where(w => !string.IsNullOrEmpty(w))
12. );
13. }
```

Yukarıda verilen verilerin temizlenmesi modülü için hazırlanan yazılımın Fix() metoduyla da Render() metodundan geçen içeriklerdeki her bir kelimenin Türkçe olup olmadığının kontrolünün yapılması amacıyla kelimeler aşağıda verilen IsTurkish() metoduyla ‘Zemberek’ kütüphanesine sokulur.

Zemberek kütüphanesinden Türkçe değil olarak dönen kelimeler için Zemberek'in sunduğu öneri sisteminden ilk önerilen kelimeyle içerik değiştirilerek kelime yazım düzeltilmesi sağlanır.

IsTurkish() metodu

```
1. public bool IsTurkish(string kelime) {  
2.     try {  
3.         if (kelime.StartsWith("@")) {  
4.             return true;  
5.         }  
6.         return _zemberek.dilTesti(kelime) == 4;  
7.     } catch (Exception) {  
8.         return false;  
9.     }
```



Tablo.2. Örnek Temizlenen Tivitler

1	hayırlı cumalar
2	Galatasaray ın yeni hocası belli oldu
3	tirnağım kırılmış hir çıkarırım bana bulasabileceğim birini etle dusurun
4	günaydın günümü aydınlatan ve bu güne beni erdiren Rabi'me şükürler olsun hayırlı nur bu cuma olar
5	ben gibi olan herkesten sen gibi olan herkese gelsin taladro sen gibi
6	en çok ilgi gören ve satan kitaplar arasında zirvede yolculuğuna devam ediyor buda geçer yahu
7	biz umudumuzu sadece allar da bağlarız
8	çok yakında siz okuyucularımızı ve dostlarımızı karşılayacağımız mekanın çizimleri geldi
9	kalbim samsunda bedenim değil ama ruhum orada başımız sağ olsun
10	loca efendinin donunu koklayıp sırtını resulullah da yüzünü kandil ve dönen kripto himmetçiler
11	ve son çalışması özeleştir yapmanın vaktinin geldiğini gösteriyor kendi blokumu
12	mali radisson da saldırı 170 kişi rehin alındı
13	ne yazık ki
14	gurur kaynağımızsın be adam
15	her şey ama her şey eskiye kaldı vakit yok bir daha yemyeşil eylül tramvaylarına edipcansever
16	bugün rahmet bugün bereket bugün şefaath bugün himmet bugün günlerin efendisi cumamız mübarek olsun
17	fatma nur un ders çalışması gerekiyordu ama olundu @ gazi mustafa kemal devlet hastanesi in Ankara Türkiye
18	sağolun başkanım
19	ben senin kollarını vatan bilmişim
20	bre cahil siz biz değil Rabbim Müslümanlar kuran ve sünnetten ayırmasın ne ete si ne hoca efendi si açın gözünüzü
21	biz sadece frikik kullanırken geri adım atıyoruz kardeş o da
22	özgürlükler beldesi tehlikede mi
23	hani ve elice de 26 mahallede sokağa çıkma yasağı
24	hayatınıza yeni bir umut ışığı katına aşk falında olacaktır fal kutusu
25	sendeki nasıl bir özgüven ya
26	siz neden yatırım yapmaktan uzak duruyorsunuz
27	gece ful motivasyon bir sürü şey planlıyorum sabah uyandığımda yataktan ameliyatla ayırıyorlar
28	mali de rehine krizi devam eden çatışmada 2 güvenlik görevlisi öldü
29	çocuklarımıza barış dolu umut dolu bir dünya diliyor dünya çocuk hakları gününü kutluyoruz dünya çocuklarıdır

4.3 Verilerin Hazırlanması

Tivitlerin ifade ettiği duyguları anlamak için duygular dört farklı sınıfta kategorize edilmiştir. Bu sınıflar ‘Mutluluk’, ‘Kızgınlık’, ‘Üzüntü’ ve ‘Şaşkınlık’ gruplarında toplanmıştır. Öğrenme algoritmalarının çalışabilmesi için tivitlerin bu dört kategori için önceden işaretlenmesi, yani bir kontrol grubunun oluşturulması gerekmektedir. Bu amaçla rastgele şekilde seçilen 62347 adet tivit gönüllülerin yardımıyla bu dört gruptan uygun görülen seçilerek, duygu işaretlenmesi yapılmıştır.

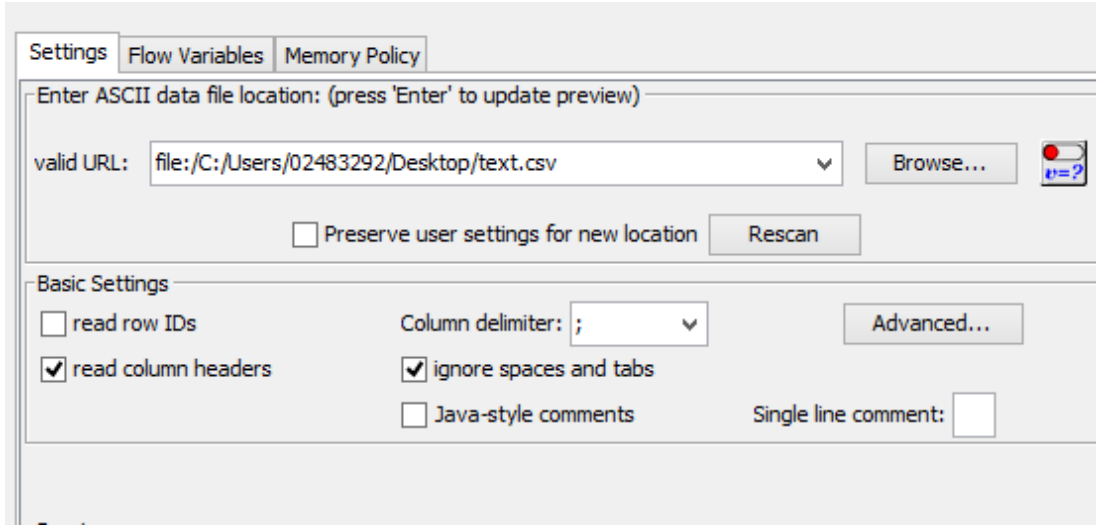
Duygu işaretlemesi yapılması sürecinde gönüllülerle Twitter üzerinden çekiliş olan bütün tivitler paylaşılmıştır. Paylaşılan tivitlerden 1 ay süre içerisinde yapabildikleri kadar duygu etiketlemesi yapmaları istenmiştir. 1 ay sürecinin sonunda toplamdaki 100 adet gönüllünün her birinden ortalama 1500 adet tivitın duyguları etiketlenmiş olarak veriler toplanmıştır. Gönüllülere her bir tivit için birden fazla duygu işaretleyebilme ya da hiçbir duygu işaretlememe (nötr duygu ifade etmesi ihtimaline karşı) konusunda esneklik sağlanmıştır. Toplanan sonuçlar da işlenirken bir tivitın belirttiği duygunun işaretlenmesi için en az 2 gönüllü tarafından aynı şekilde işaretlenmiş olması kriteri kıstas alınmıştır.

Tablo.3. Örnek etiketlenen tivitler

	Text	Mutlu	Kızgın	Şaşkın	Üzgün
1	hayırlı cumalar	1	0	0	0
2	Galatasaray ın yeni hocası belli oldu	0	0	0	0
3	tirnağım kırılmış hir cıkarırım bana bulasabileceğim birini etle dusurun	0	1	0	0
4	günaydın günümü aydınlatan ve bu güne beni erdiren Rabi'me şükürler olsun hayırlı nur bu cuma olar	1	0	0	0
5	ben gibi olan herkesten sen gibi olan herkese gelsin taladro sen gibi	0	0	0	1
6	en çok ilgi gören ve satan kitaplar arasında zirvede yolculuğuna devam ediyor buda geçer yahu	0	0	0	1
7	biz umudumuzu sadece allar da bağlarız	0	1	0	0
8	çok yakında siz okuyucularımızı ve dostlarımızı karşılayacağımız mekanın çizimleri geldi	1	0	0	0
9	kalbim samsunda bedenim değil ama ruhum orada başımız sağ olsun	0	0	0	1
10	loca efendinin donunu koklayıp sırtını resulullah da yüzünü kandil ve dönen kripto himmetçiler	0	0	0	1
11	ve son çalışması özeleştir yapmanın vaktinin geldiğini gösteriyor kendi blokumu	0	1	0	0
12	mali radisson da saldırı 170 kişi rehin alındı	0	1	0	0
13	ne yazık ki	0	0	0	1
14	gurur kaynağımızsın be adam	1	0	0	0
15	her şey ama her şey eskiye kaldı vakit yok bir daha yemyeşil eylül tramvaylarına edipcansever	0	0	0	1
16	bugün rahmet bugün bereket bugün şefaath bugün himmet bugün günlerin efendisi cumamız mübarek olsun	1	0	0	0
17	Fatma nur un ders çalışması gerekiyordu ama olundu	0	0	1	0

	@ gazi mutafa kemal devlet hastanesi in anlara türküye				
18	sağolun başkanım	1	0	0	0
19	ben senin kollarını vatan bilmişim	0	0	1	0
20	bre cahil siz biz değil Rabi'm Müslümanlar kuran ve sünnetten ayırmasın ne ete si ne hoca efendi si açın gözünüzü	0	1	0	0
21	biz sadece frikik kullanırken geri adım atıyoruz kardeş o da	0	1	0	0
22	özgürlükler beldesi tehlikede mi	0	0	1	0
23	hani ve elice de 26 mahallede sokağa çıkma yasağı	0	0	1	0
24	hayatınıza yeni bir umut ışığı katına aşk falında olacaktır fal kutusu	1	0	0	0
25	sendeki nasıl bir özgüven ya	0	0	1	0
26	siz neden yatırım yapmaktan uzak duruyorsunuz	0	0	1	0
27	gece ful motivasyon bir sürü şey planlıyorum sabah uyandığımda yataktan ameliyatla ayırıyorlar	0	0	0	1
28	mali de rehine krizi devam eden çatışmada 2 güvenlik görevlisi öldü	0	0	0	1
29	çocuklarımıza barış dolu umut dolu bir dünya diliyor dünya çocuk hakları gününü kutluyoruz dünya çocuklarıdır	1	0	0	0
30	namaz kılan bir toplumun psikolojiye tek t veren bir toplumun da sosyolojiye ihtiyacı yoktur cemil meri protesto ediyor	0	1	0	0

File Reader: Verilerin Knime üzerinde işlenebilmesi için DocumentDB üzerinde tutulan veriler CSV formatına çevrilmiştir.



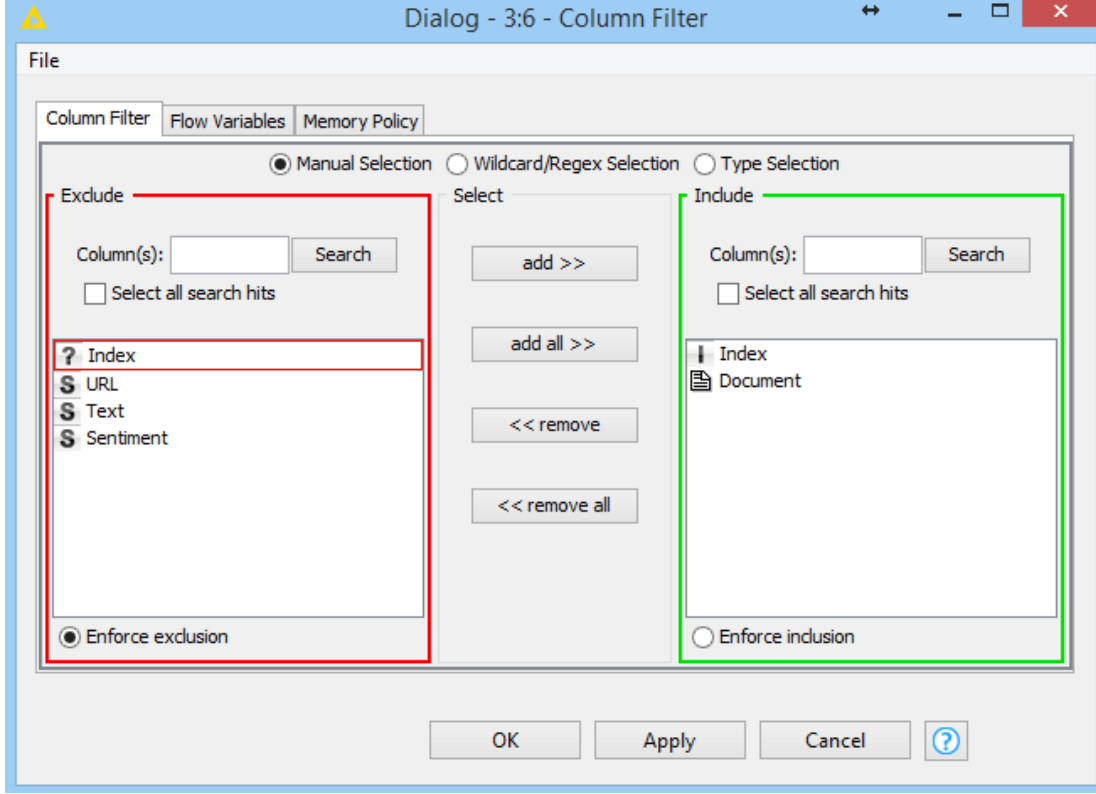
Şekil.11. Knime programında kullanılan File Reader modülü konfigürasyonu

Strings To Document: Gereksiz içeriklerin kullanılmaması amacıyla veriler gruplandırılmıştır.

Row ID	Index	URL	Text	Sentiment
Row0	1001	http://www.burcinadak.com/url/comment/1	kendinize değer verin	Mutlu
Row1	1002	http://www.burcinadak.com/url/comment/2	yaşasın sonunda geldi	Mutlu
Row2	1003	http://www.burcinadak.com/url/comment/3	eteklerim zil çalıyor	Mutlu
Row3	1004	http://www.burcinadak.com/url/comment/4	sabırsızlıkla bekliyorum	Mutlu
Row4	1005	http://www.burcinadak.com/url/comment/5	lütfen hiç bitmesin	Mutlu
Row5	1006	http://www.burcinadak.com/url/comment/6	seni veren allaha ş...	Mutlu
Row6	1007	http://www.burcinadak.com/url/comment/7	bir an önce yarın ol...	Mutlu
Row7	1008	http://www.burcinadak.com/url/comment/8	sonunda güneş doğ...	Mutlu
Row8	1009	http://www.burcinadak.com/url/comment/9	hoş geldin bahar	Mutlu
Row9	1010	http://www.burcinadak.com/url/comment...	güzel kuşlar size de...	Mutlu
Row10	1011	http://www.burcinadak.com/url/comment...	yeşillikler içinde huz...	Mutlu
Row11	1012	http://www.burcinadak.com/url/comment...	daha ne isteyebilir...	Mutlu
Row12	1013	http://www.burcinadak.com/url/comment...	çok tatlısın çok	Mutlu
Row13	1014	http://www.burcinadak.com/url/comment...	günaydın güzellikler	Mutlu
Row14	1015	http://www.burcinadak.com/url/comment...	mavi gözlü sevgilimi...	Mutlu
Row15	1016	http://www.burcinadak.com/url/comment...	hoş geldin tatil	Mutlu
Row16	1017	http://www.burcinadak.com/url/comment...	işte beklenen an geldi	Mutlu
Row17	1018	http://www.burcinadak.com/url/comment...	heyecan dorukta	Mutlu

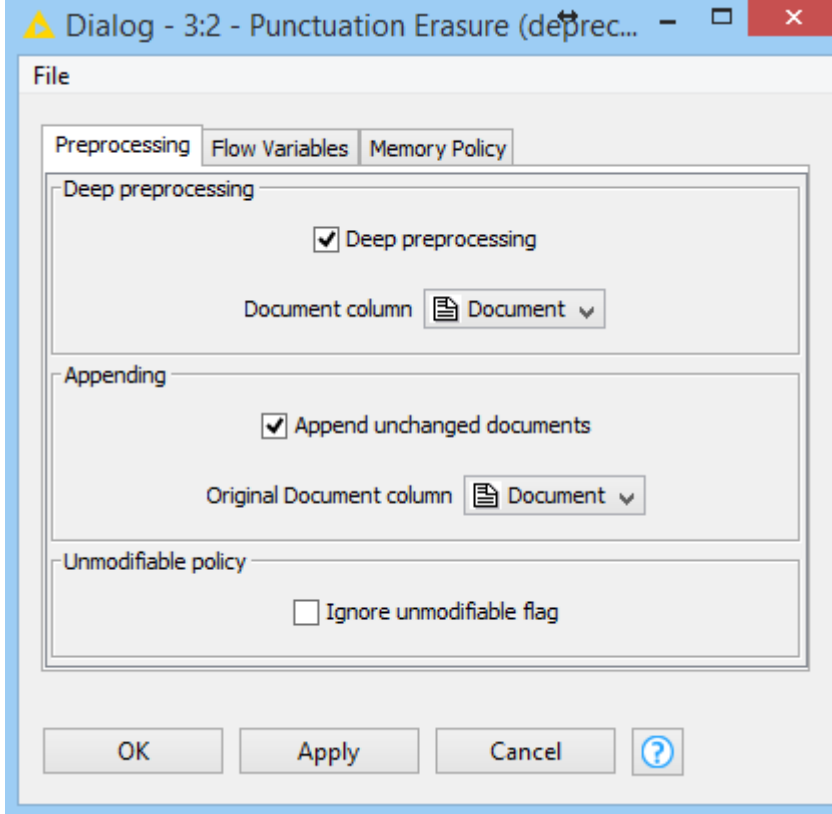
Şekil.12. Knime programında kullanılan String To Document modülü konfigürasyonu

Column Filter: Document altında toplanan veriler Column Filter ile filtrelenmiştir. Bu sayede diğer gruplarda toplanan kullanılmasına gerek duyulmayan veriler çıkartılmıştır.



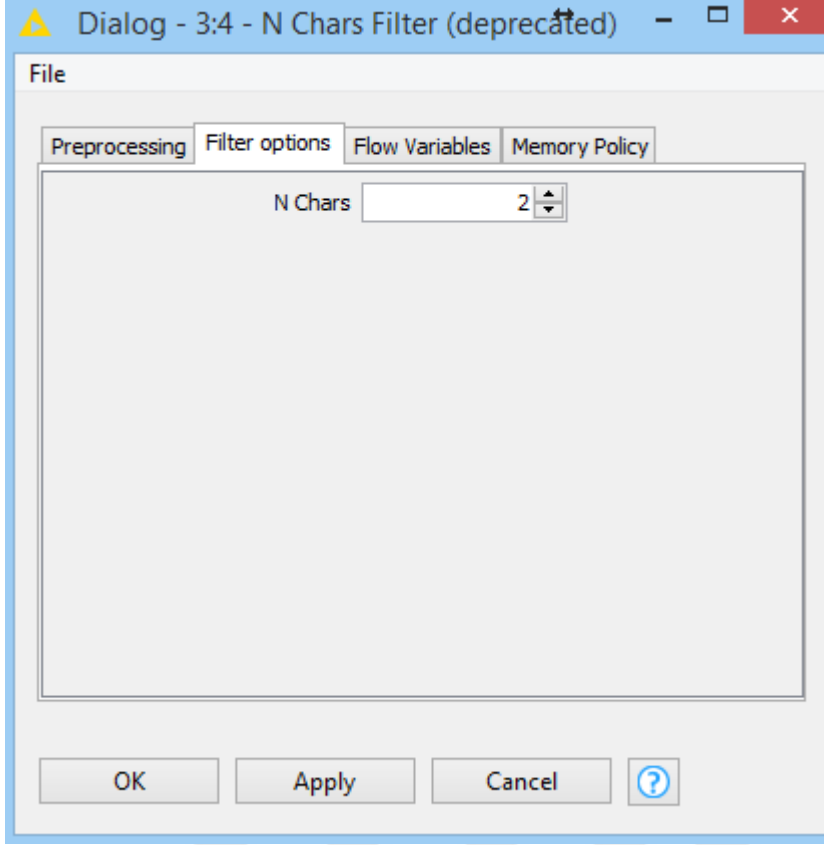
Şekil.13. Knime programında kullanılan Column Filter modülü konfigürasyonu

Punctuation Erasure: Noktalama işaretleri ve özel karakterler Punctuation Erasure kullanılarak verilerden çıkartılmıştır. Veri temizleme modülünde zaten noktalama işaretleri ve özel karakterlerin çıkartılıyor olmasına rağmen ilerde ayrıca işlenmemiş bir verinin de programa analiz edilmesi için eklenebileceği öngörülerek koyulmuştur.



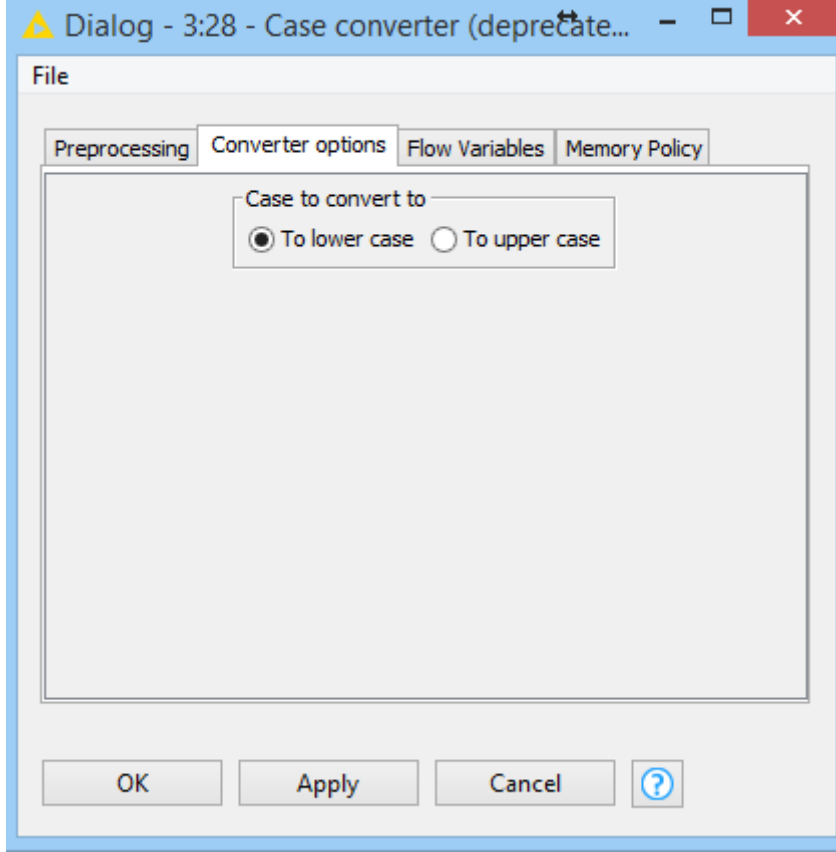
Şekil.14. Knime programında kullanılan Punctuation Erasure modülü konfigürasyonu

N Chars Filter: 2 harf ve 1 harften oluşan kelimeler her hangi bir duygu ifade etmeyeceği öngörülerek çıkartılmıştır. Yani, en az 3 harften oluşan kelimeler analiz edilmesi için tutulmuştur.



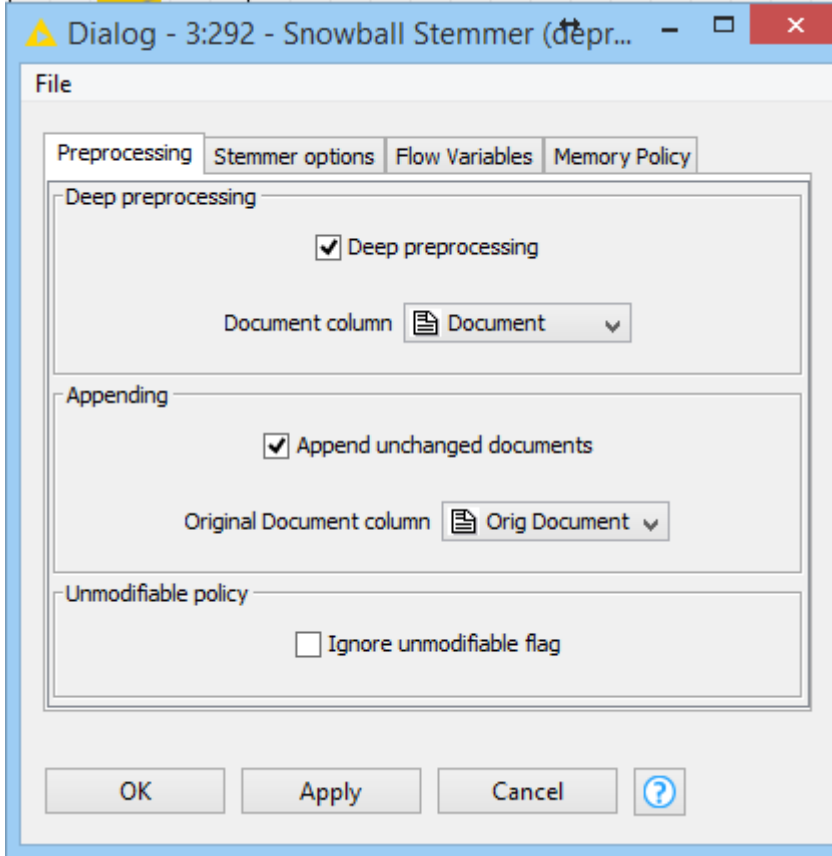
Şekil.15. Knime programında kullanılan N Chars Filter modülü konfigürasyonu

Case Converter: Öğrenme algoritmalarının büyük harf ve küçük harfle başlayan iki aynı kelime farklı algılanmasını engellemek için Case Converter yardımı ile bütün kelimeler küçük harfe çevrilmiştir. Veri temizleme modülünde zaten verilerin hepsinin küçük harfe otomatik olarak çevriliyor olmasına rağmen ilerde ayrıca işlenmemiş bir verinin de programa analiz edilmesi için eklenebileceği öngörülerek koyulmuştur.



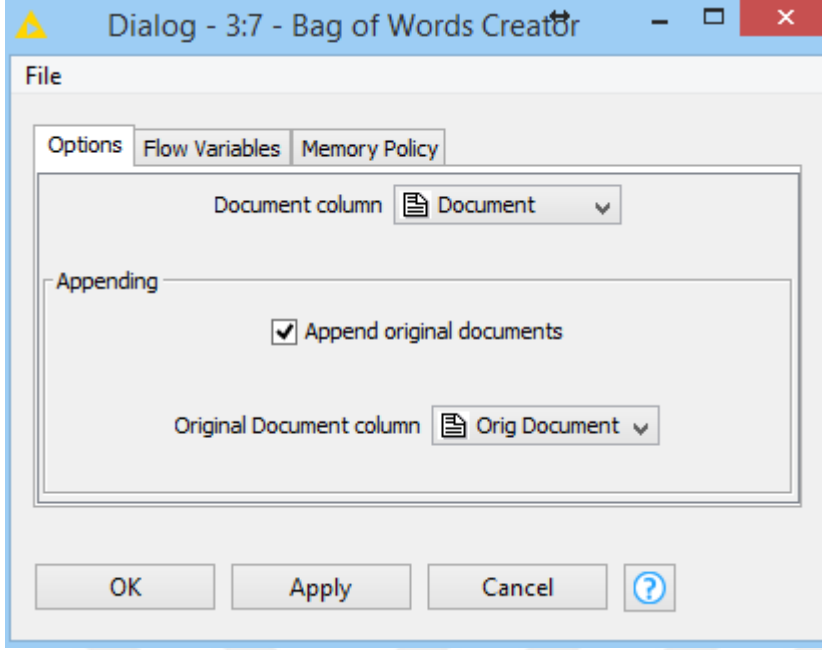
Şekil.16. Knime programında kullanılan Case Converter modülü konfigürasyonu

Snowball Stemmer: Snowball Stemmer ile her bir kelimenin köklere arılması sağlanmıştır.



Şekil.17. Knime programında kullanılan Snowball Stemmer modülü konfigürasyonu

Bag Of Words Creator: Snowball Stemmer ile kök haline indirgenmiş kelimeler toplanır.

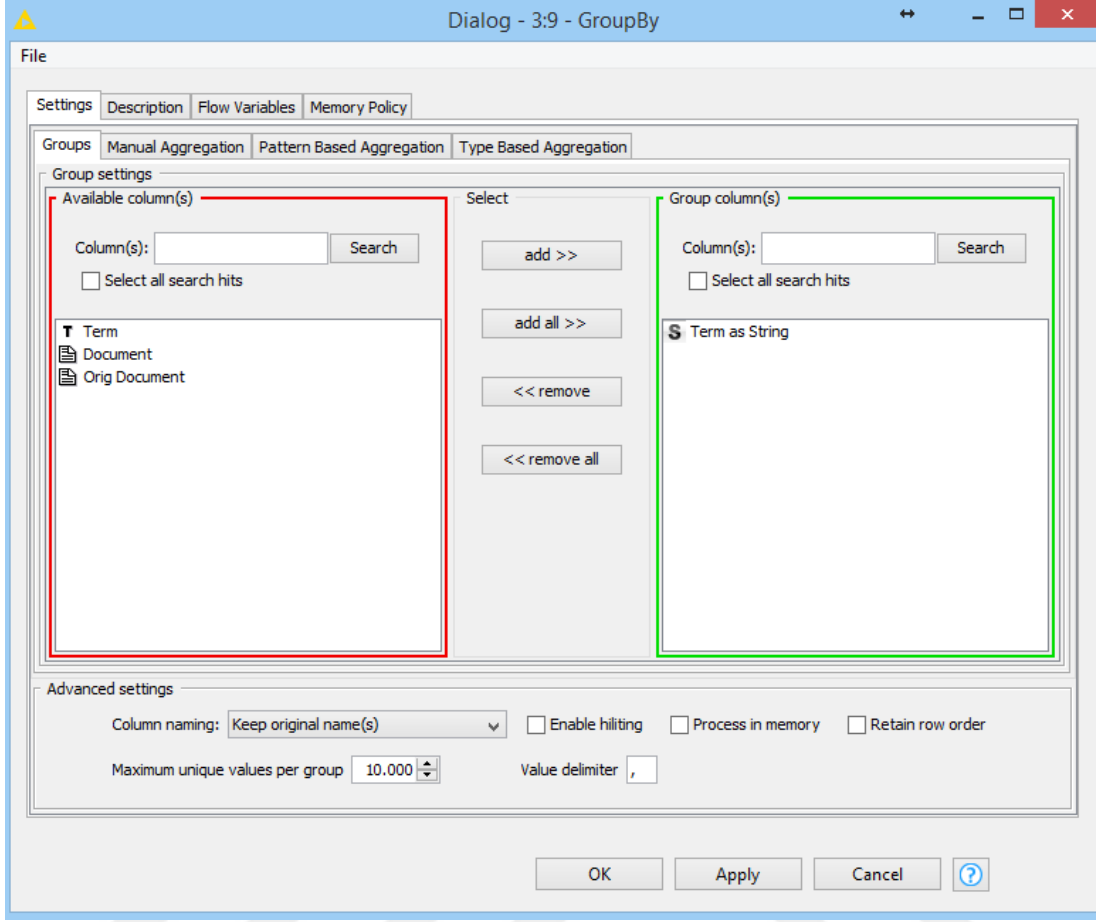


Şekil.18. Knime programında kullanılan Bag Of Words Creator modülü konfigürasyonu

Term To String: Terimler işlenebilmesi amacıyla string formatına dönüştürülmüştür.

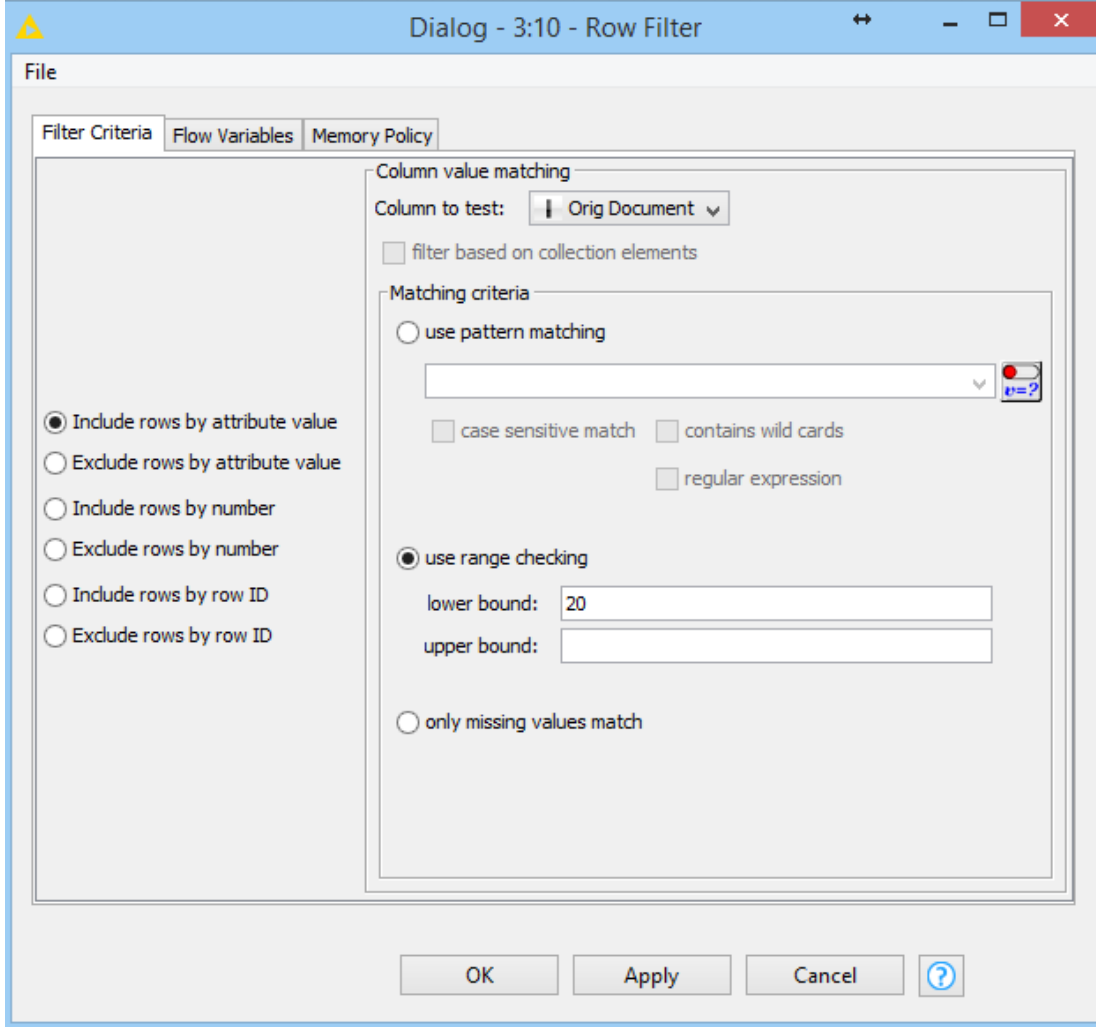
Term Frequency: Kelimenin tivit içeriğinde kullanılma sıklığını ölçmek için kullanılmıştır.

Group By: Toplam veriler içerisinde her bir kelimenin kullanılma sıklığına göre kelimeleri gruplar.



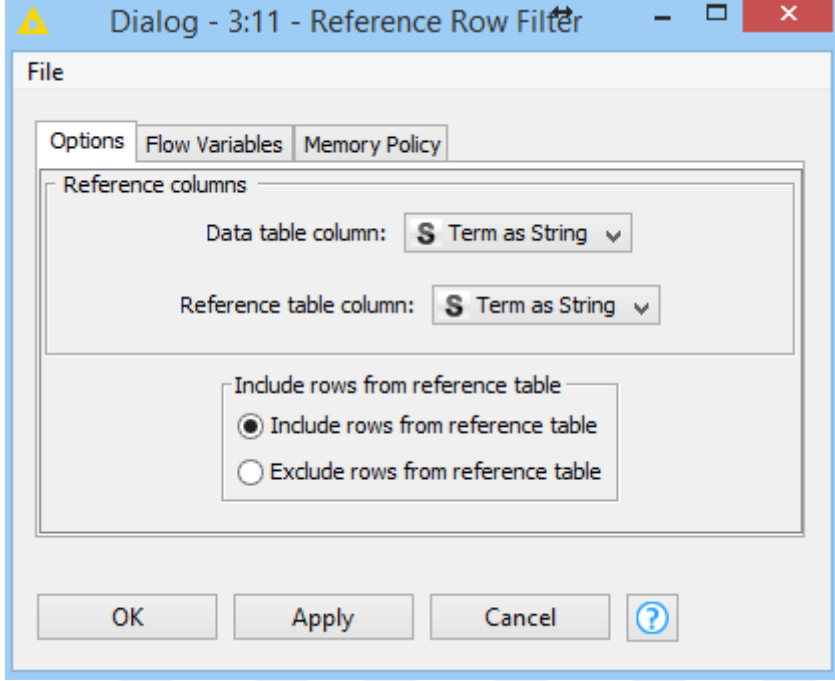
Şekil.19. Knime programında kullanılan Group By modülü konfigürasyonu

Row Filter: Belirli bir adetin altında kullanılan kelimeler öğrenme algoritmalarını olumsuz yönde etkileyecek ve bir desen bulunmasını zorlaştırırlar. Bu nedenle toplanan tivitler arasında 20 âdetin altında kullanılmış olan kelimelerin olduğu tivitlerin dikkate alınmaması gerekmektedir. Knime üzerinde “Row filter” özelliği kullanılarak bu tivitler filtrelenmiştir.



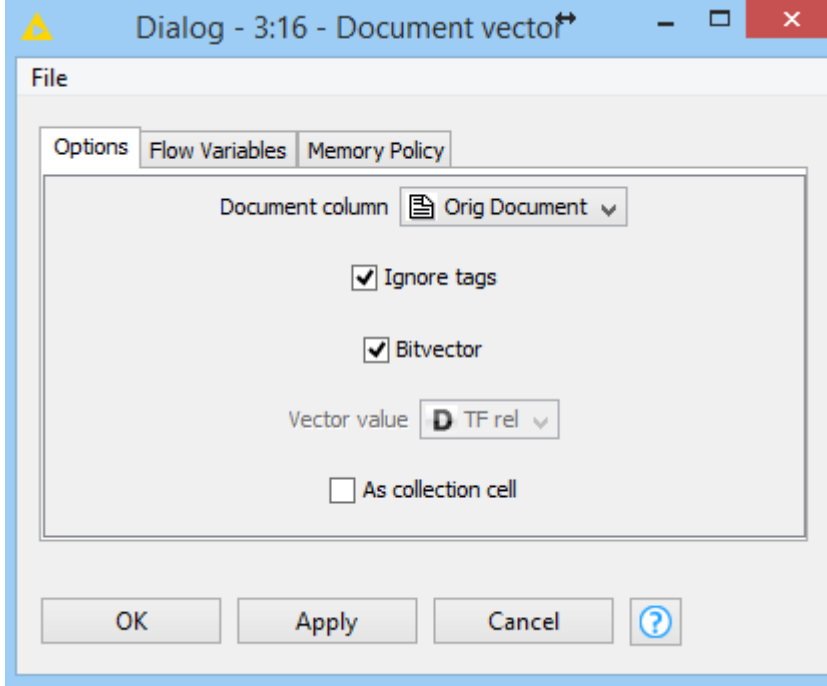
Şekil.20. Knime programında kullanılan Row Filter modülü konfigürasyonu

Reference Row Filter: Kullanım sıklığına göre ayrılarak filtreleniş kelimelere göre bag of words içeriğinde toplanan verileri filtreler.



Şekil.21. Knime programında kullanılan Reference Row Filter modülü konfigürasyonu

Document Vector: Her bir tivit içeriğinin her bir kelimeye göre kullanılma oranını gösteren matris yapısını oluşturur.

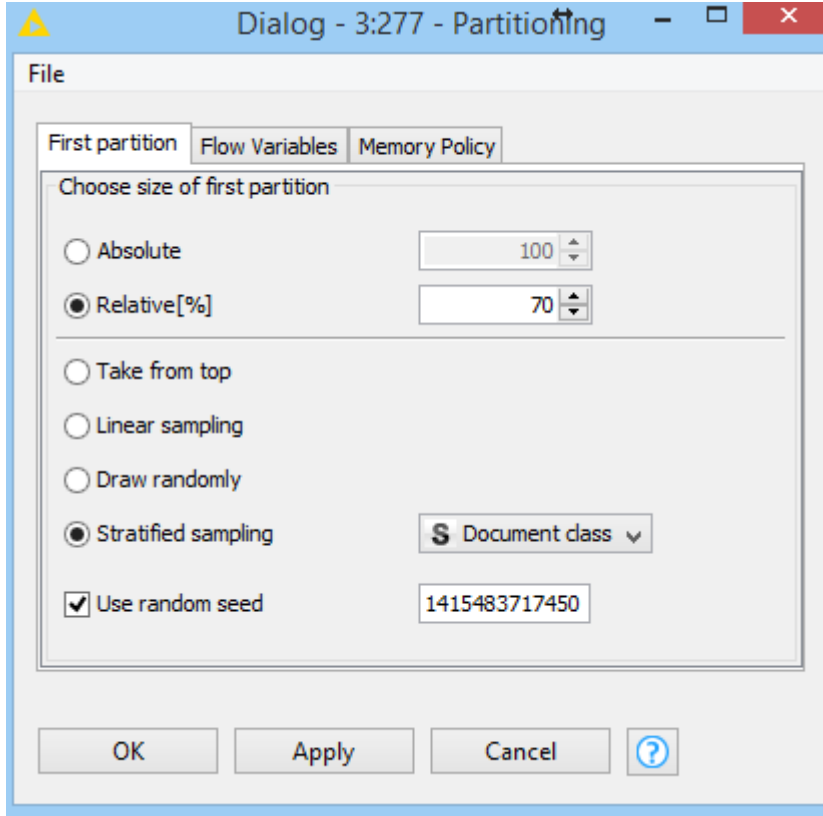


Şekil.22. Knime programında kullanılan Document Vector modülü konfigürasyonu

Category to Class: Her bir tivit içeriğinin her bir kelimeye oranla kullanılma içeriğini gösteren matrisin her bir satırına ilgili duyguyu ekler.

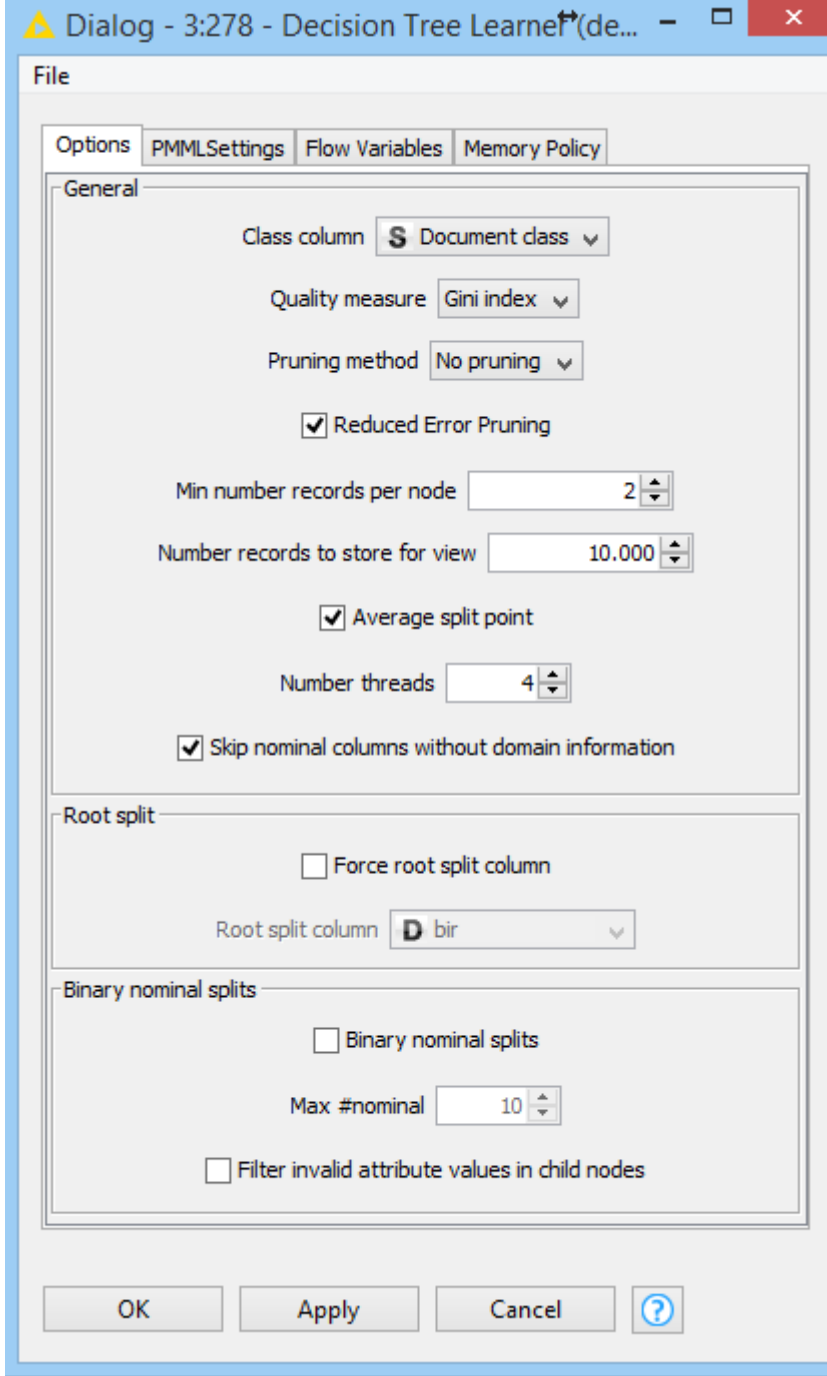
Color Manager: Elde edilen verilerin gösteriminde sadece görsel olarak kolay ayırım yapılmasını sağlaması amacıyla kullanılmıştır.

Partitioning: Partitioning ile verilerin %70'i öğrenme yani kontrol grubuna, kalan %30'u da test yani deney grubuna ayrılmıştır.



Şekil.23. Knime programında kullanılan Partitioning modülü konfigürasyonu

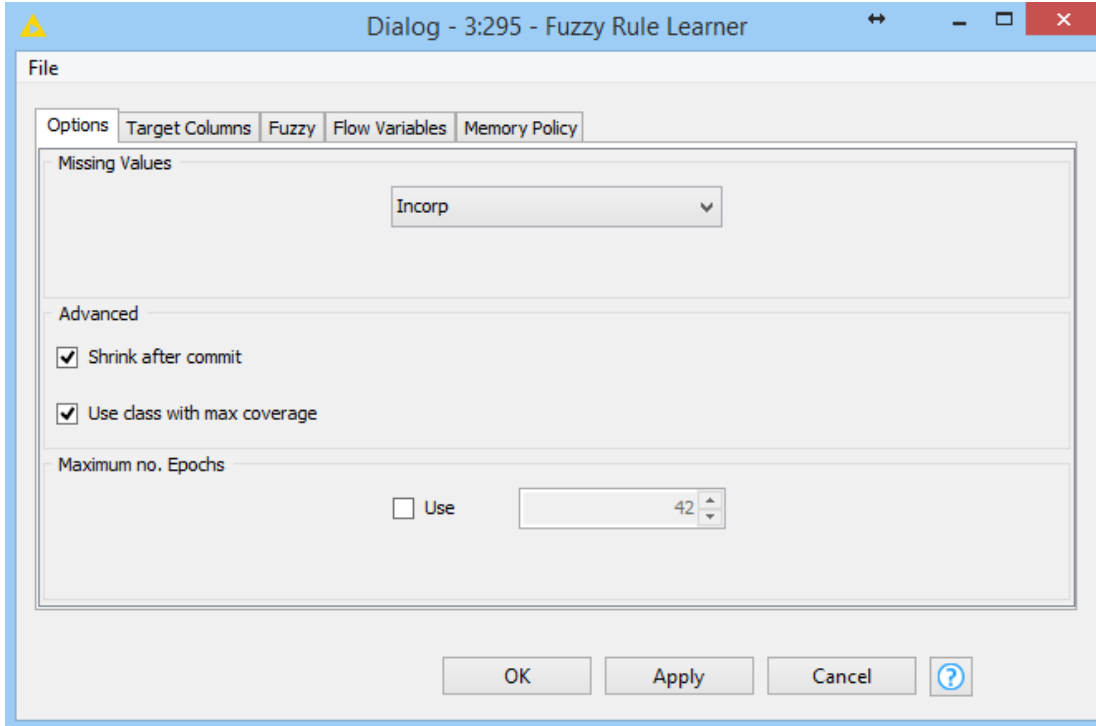
Decision Tree Learner: Decision Tree Learner ile veriler Karar Ağacı analizine sokulmuştur.



Şekil.24. Knime programında kullanılan Decision Tree Learner modülü konfigürasyonu

Decision Tree Predictor: Decision Tree Predictor ile, Decision Tree Learner'a sokulan verilerin duygu tahminlemesi sağlanmıştır.

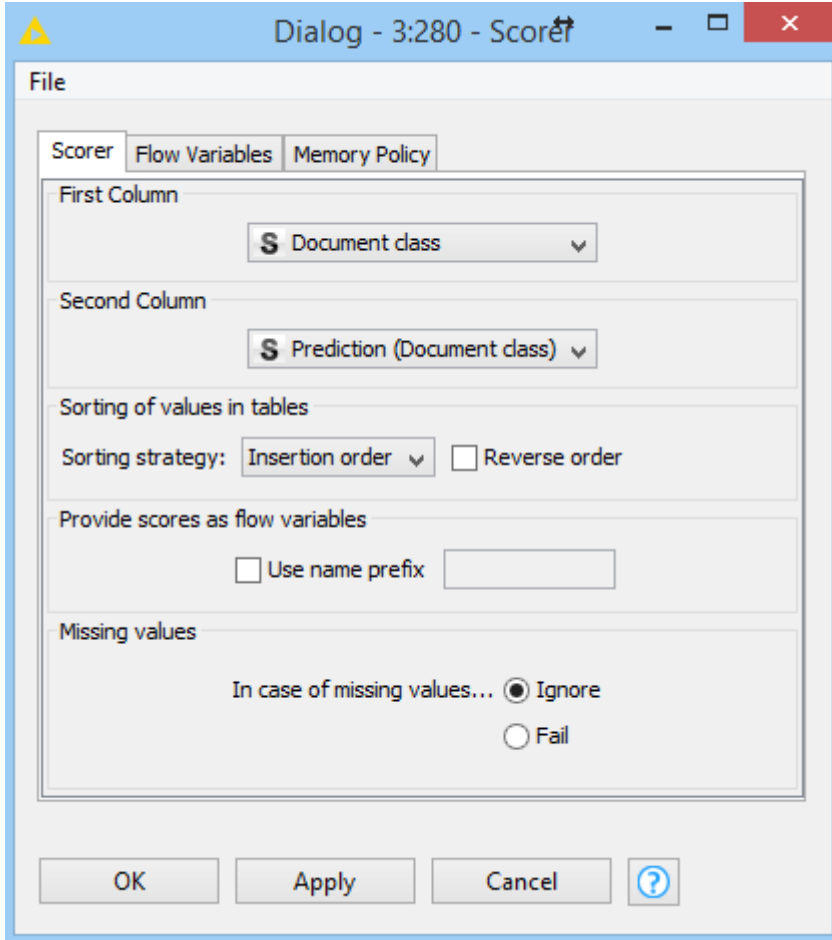
Fuzzy Rule Learner: Fuzzy Rule Learner ile veriler Bulanık Mantık analizine sokulmuştur.



Şekil.25. Knime programında kullanılan Fuzzy Rule Learner modülü konfigürasyonu

Fuzzy Rule Predictor: Fuzzy Rule Predictor ile, Fuzzy Rule Learner'a sokulan verilerin tahminlenmesi sağlanmıştır.

Scorer: Scorer ile Karar Ağacı ve Bulanık Mantık analizine sokulan verilerin sonuçları alınmıştır.



Şekil.26. Knime programında kullanılan Scorer modülü konfigürasyonu

5. SONUÇ

Tez kapsamında Türkçe içerikli tivitlerin duygu analizinin yapılabilmesi amaçlanmıştır. Bu doğrultuda, eğitilen sistem tarafından, her bir elde edilen tivit ‘Mutlu, Üzgün, Kızgın, Şaşkın’ duygularından birisi ile etiketlenmesi sağlanmıştır. Kontrol grubu verilerinden yararlanılarak deney grubundan elde edilen sonuçlar Karar ağacı ve bulanık kural yöntemleriyle ayrı ayrı değerlendirilmiştir.

Karar ağacı ile aşağıdaki skorlar elde edilmiştir. Çalışma sonucunda %85,372’lik başarı oranı yakalanmıştır.

Tablo.4. Karar Ağacı ile Duygu Analizi Sonuçları

Duygu/Tahmin	Mutlu	Üzgün	Kızgın	Şaşkın
Mutlu	2910	350	1526	0
Üzgün	285	1617	536	0
Kızgın	100	270	4781	87
Şaşkın	88	356	737	535

Fuzzy Rule Learner kullanılarak bulanık kural yöntemi ile aşağıdaki skorlar elde edilmiştir. Çalışma sonucunda %83,608’lik başarı oranı yakalanmıştır.

Tablo.5. Fuzzy Rule Learner ile Duygu Analizi Sonuçları

Duygu/Tahmin	Mutlu	Üzgün	Kızgın	Şaşkın
Mutlu	4786	0	0	0
Üzgün	342	2096	0	0
Kızgın	1067	196	3975	0
Şaşkın	557	0	0	997

6. DEĞERLENDİRME

Elde edilen sonuçlar doğrultusunda bulanık kural ile karar ağacı yöntemlerinin çok yakın başarı yüzdeleri bulunmasına rağmen duygu kategorilendirme konusunda iki yöntemin de farklı kategoriler için güven veremeyecek sonuçlar verdiği görülmüştür. Buna rağmen kategorilerde yakından inceleme yapıldığında bulanık karar yönteminin bu çalışmada daha geçerli bir yöntem olduğu aşikârdır.

Başka bir açıdan bakıldığında ise yöntemlerin ikisinin de en güvenilir sonuçları ‘Üzgün’ kategorisinde verirken, ‘Kızgın’ ve ‘Mutlu’ kategorilerini karıştırdığı görülmüştür. Buna istinaden Türkçe’de üzüntü belirten kelime ve cümle yapılarının daha kesin, ancak kızgınlık ve/veya mutluluk ifade eden kelime ve cümle yapılarının daha açık uçlu olduğu söylenebilir.

Açık uçlu kelime ve cümle yapılarının duygu kategorilendirme çalışmalarında mecazî anlam sözlüğü ya da duygu sözlüğü kullanılarak daha kesin sonuçlar verebilecektir.

Ayrıca verilerin cümle öğelerine ayrılarak kelimenin cümle içinde anlam/ifade yoğunluğunun tespiti için ağırlıklandırma kullanılarak tezin daha kesin sonuçlara ulaşması sağlanabilir.

Elde edilen sonuçlarda Türkçe kelimeler yakalanırken özel isimlerin yabancı isimler gibi algılandığı da görüldü. Örneğin, ‘Fatma’ ismi Tivit temizleme modülüne sokulduğunda ‘satma’ olarak çevrildi. Diğer bir örnek de şehirlerden verilirse ‘Ankara’ veri temizleme modülüne sokulduğunda ‘anlara’ olarak çevrildiği görülmüştür. Bu nedenle isimlerin Türkçe olup olmadığına bakmak için Zemberek kütüphanesine sokulmadan önce özel isimler zaten duygu ifade etmeyeceklerinden ilgili sözlüklere sokularak elenmeliydi.

Tablo.6. Analiz edildiğinde hatalı dönen örnek duygu ifadeleri

Text	Olması gereken duygu	Tahmin edilen duygu
kutlu doğum ayı oldu o artık	Şaşkın	Mutlu
demek cuma ya gitmiyorsun	Şaşkın	Kızgın
aradığınız insanlığa şuan ulaşamıyor ama siz yine de arasına arayın	Kızgın	Şaşkın
evet bir siirde yerini kaybetmiş bir cumleyim karanlığın sonsuzluğunda acılar canımı alırcasına yükleniyorlar sırtıma	Üzgün	Kızgın
sıradan geçen günlerimin belirsiz zaman dilimlerinde çok güzel bir şeyi kaybettiğim hissi ile yüz yüze gelmekten yoruldum	Üzgün	Şaşkın
300 Ispartalı ya 2500 yıl boyunca unutmayan tarih bugünün zalimlerini de zulme onurluca direnenleri de unutmuyacaktı	Kızgın	Şaşkın
dualarımız ve desteklerimiz kardeşlerimize allar yardımcıları olsun türkmendağındakatliamvar	Üzgün	Kızgın
oysa ben dua varla bekleyip sevmistim seni neden bilebilirdim ki gelmeden cekip gideceğini	Üzgün	Kızgın
kıskandırıcım derken kendinizden soğutuyorsunuz yapmayın	Kızgın	Şaşkın
facebook kimsenin babasının tapulu malı değil	Kızgın	Şaşkın
adamım	Mutlu	Şaşkın
mücadelem buna yetmez	Üzgün	Şaşkın
savaşta hayatını kaybedenlerin 25inin	Şaşkın	Üzgün

çocuk olduğunu biliyor muydunuz		
------------------------------------	--	--



KAYNAKLAR

[1] Binali, H., Wu, C., Potdar, V. (2010). Computational Approaches for Emotion Detection in Text.

https://www.researchgate.net/publication/224185404_Computational_approaches_for_emotion_detection_in_text

[2] Meral, M. (2014). Twitter Üzerinde Duygu Analizi.

<http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel7%2F6820096%2F6830164%2F06830323.pdf>

[3] Botego. (2012). Sentiment Analizi

<http://www.sentimentanalizi.com/>

[4] Ucan, A. (15.10.2014). Otomatik Duygu Analizi Sözlüğü ve Duygu Analizinde Kullanılması.

<http://www.openaccess.hacettepe.edu.tr:8080/xmlui/bitstream/handle/11655/2626/100811e6-ace7-4ded-972e-06577a8fd46f.pdf?sequence=1&isAllowed=y>

[5] Lee, C., Lee G. G. (2011). Emotion Recognition for Affective User Interfaces using Natural Language Dialogs

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.127.1808&rep=rep1&type=pdf>

[6] Cervenec, R., Burget, R. (20.09.2011). Identifying Expression of Emotions in Czech Text Using Semantic Relations for Dimension Reduction

<http://elektrorevue.cz/en/articles/analogue-technics/0/identifying-expression-of-emotions-in-czech-text-using-semantic-relations-for-dimension-reduction/>

[7] Zhong, Y., Guo, Z. (14.12.2015). OMG! How do you know my feeling? A predictive model to determine sentiments from Tweets

<https://cseweb.ucsd.edu/~jmcauley/cse190/reports/sp15/026.pdf>

[8] Vikipedi. "Twitter" <https://tr.wikipedia.org/wiki/Twitter> 22 Ağustos 2015

- [9] Twitter Inc. “Products and Services, API Overview”
<https://dev.twitter.com/overview/api> 2016
- [10] Hammer, E. “About OAuth” <https://oauth.net/about/> 5 Eylül 2007
- [11] Wikipedi. (01.08.2016) “JSON” <https://tr.wikipedia.org/wiki/JSON>
- [12] Mayo, Joe. (26.04.2015) “Linq2Twitter”
<https://linqtotwitter.codeplex.com/>
- [13] (01.09.2016) "NoSQL Relational Database Management System".
<http://www.strozzi.it/>
- [14] Wikipedia.(21.01.2017) “NoSQL”. <https://en.wikipedia.org/wiki/NoSQL>
- [15] (01.10.2016) DocumentDB'ye giriş: Bir NoSQL JSON Veritabanı
“DocumentDB verileri nasıl yönetir?” <https://azure.microsoft.com/tr-tr/documentation/articles/documentdb-introduction/>
- [16] Wikipedi. (22.01.2017) “Zemberek (yazılım)”
[https://tr.wikipedia.org/wiki/Zemberek_\(yazılım\)](https://tr.wikipedia.org/wiki/Zemberek_(yazılım))
- [17] Wikipedia. (10.12.2016) “Knime” <https://en.wikipedia.org/wiki/KNIME>
- [18] Wikipedi. (02.10.2016) “Bulanık Mantık”
https://tr.wikipedia.org/wiki/Bulan%C4%B1k_mant%C4%B1k
- [19] Wikipedi. “Veri Madenciliği”
https://tr.wikipedia.org/wiki/Veri_madencili%C4%9Fi (21.01.2017)

ÖZGEÇMİŞ

BURCIN ADAK KAPLAN

E-Posta : burcinadak@hotmail.com

Adres Bilgileri : Türkiye - İstanbul(Asya) - Ataşehir - Küçükbakkalköy

Cep Telefonu : 90 (555) 717 00 41

Doğum Tarihi : 01.06.1986

EĞİTİM

(2007-2012) Orta Doğu Teknik Üniversitesi (ODTÜ)– Fizik
(Bitirme Tezi: Bilgisayar ile Spektrometre Arasında Arayüz Oluşturmak)

(2004-2007) İstanbul Aydın Üniversitesi (IAU)– Bilgisayar Teknolojileri ve Programlama (Bitirme Tezi: ASP.Net tabanlı Web Sitesi)

(2002-2004) Bingül Erdem Lisesi Fen-Matematik

(2000-2002) Nişantaşı Nuri Akın Anadolu Lisesi

(1995-2000) Şair Behçet Kemal Çağlar İlk Öğretim Okulu

TECRÜBE

(2016-Halen) Sqills Bilgi Teknolojileri

S3 Passenger: Kara ve Raylı seyahat firmalarının rezervasyon, bilet alım, iptal işlemlerini yönettikleri online platformdur..

S3Passenger sisteminde ortaya çıkan müşteri ihtiyaçları doğrultusunda teknik analizler yaparak yapılması uygun görülen ihtiyaçların teknik ekibe aktarımı sürecinde rol almaktayım.

(2015-2016) Formalis Bilgi Teknolojileri

Mi4biz: firmaların müşteri şikayet ve taleplerini toplayabildiği ve yönetebildiği online platform.

Mi4biz sisteminde entegrasyonların ve yeni geliştirmelerin süreçlerinin hazırlanmasında, uygulanması ve dokümanite edilmesinde görev almaktayım. Ayrıca Selenium ile UI testlerinin otomatize edilmesini sağlayan arayüzün geliştirmesini (C#, Selenium Web Driver) yaptım. Selenium ile oluşturmuş olduğum arayüzün geliştirmelerine de devam etmekteyim.

(2013-2015) Kariyer.net Ar-Ge - İş Analisti

MemnuniyetAnalizi.net (www.memnuniyetanalizi.net) : Firmaların çalışan memnuniyet anketlerini ölçme ve değerlendirmelerini sağlayan online platform Tüm İşler Burada (www.tumislerburada.com): İnternet üzerinde yayınlanan iş ilanlarının hepsini tek bir çatı altında toplayarak kullanıcıların tek bir yerden hepsine ulaşmalarını sağlayan web sitesi KariyerLink v.3 : Firmaların web sitelerine entegre edilen, iş ilanlarını yayınladıkları uygulama İK Sayfam (www.iksayfam.com) : KariyerLink'in ücretsiz olan versiyonudur. Firmalara iş ilanlarını deneme amaçlı olarak KariyerLink'e göre daha kısıtlı özelliklerle ücretsiz olarak web sitelerinde yayınlamalarını sağlar Ayrıca Ar-Ge departmanımız çatısında yürütülen iki ayrı Ar-Ge projesinde de görev aldım

(2012-2015) Sensetek AB – Kurucu Müdür

High Sensitive Master/Slave Robotic Arm projemin Bilim, Sanayi ve Teknoloji Bakanlığı'nın Teknogirişim desteğini almasıyla kurulmuş olan şirkettir.

High Sensitive Master/Slave Robotic Arm: Bilgisayara bağlı olan bir robot kolun başka bir bilgisayarın başındaki kişi tarafından eş zamanlı olarak ve eş hassasiyetle hareket ettirilebildiği projedir.

(2010-2011) Çukurambar Sınav Dershanesi - Fen Bilgisi Öğretmeni (Yarı Zamanlı)

(2009-2010) Kızılay Bulvar Atak Dershanesi - Fen Bilgisi Öğretmeni (Yarı Zamanlı)

(2008-2009) Ankara Dershanesi – Fen Bilgisi Öğretmenliği (Yarı Zamanlı)

(2005-2006) İstanbul Aydın Üniversitesi- Bilgi İşlem (Staj)

(2004-2005) İstanbul Bellcom Bilgisayar - Teknik Destek (Staj)

YABANCI DİLLER

İngilizce (İleri)

BİLGİSAYAR BİLGİSİ

PROGRAMLAMA DİLLERİ :

MS Office (Word, Access, Excel, Visio, MS Project), Balsamiq

MockUps, Flash, Photoshop, Freehand. TFS, DreamWeaver

LabView, C#, HTML, CSS, SQL, SPARQL, JIRA, XML, UML

SERTİFİKALAR

PMP – PMI

Uluslararası İş Analistliği – PEM

Web Tasarımı - İSMEK

İş ve Sosyal Güvenlik Hukuku - KAVED

Başarılı Lider Olma - Psikologlar ve Psikiyatristler Derneği

Stres ve Stresle Başa Çıkma - Psikologlar ve Psikiyatristler Derneği

Beden Dili - Psikologlar ve Psikiyatristler Derneği

İnsan Kaynakları Yönetimi- KAVED

İş ve Çalışma Psikolojisi – Psikologlar ve Psikiyatristler Derneği

Kişilik Psikolojisi – Psikologlar ve Psikiyatristler Derneği

BURSLAR/BAŞARILAR

High Sensitive Master/Slave Robotic Arm projesi: 2013 senesinde Bilim, Sanayi ve Teknoloji

Bakanlığı'nın Teknogirişim desteği ile Sensetek AB Ar-Ge Mühendislik şirketinin kurulma-

sıyla hayata geçirilen bir projedir. Proje yine 2013 senesinde Intel Global Challenge

Turkey'de finallere kalıp plaket sahibi olmuştur.

HOBİLER

Yoga, seyahat

ÜYE OLUNAN TOPLULUKLAR

İstanbul ODTÜ Mezunlar Derneği