

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

METİN MADENCİLİĞİ İLE SOSYAL MEDYA ANALİZİ
YÜKSEK LİSANS TEZİ

Hazırlayan
Gökhan ARAVİ

Tez Danışmanı
Yrd. Doç. Dr. Metin ZONTUL

İSTANBUL-2014

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

METİN MADENCİLİĞİ İLE SOSYAL MEDYA ANALİZİ
YÜKSEK LİSANS TEZİ

Hazırlayan
Gökhan ARAVİ

Tez Danışmanı
Yrd. Doç. Dr. Metin ZONTUL

İSTANBUL-2014



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ
MÜDÜRLÜĞÜ'NE

Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Mühendisliği (Tezli) Yüksek Lisans Programı Y1113.010014 numaralı öğrencisi **Gökhan ARAVİ**' nin "**METİN MADENCİLİĞİ İLE SOSYAL MEDYA ANALİZİ**" adlı tez çalışması Enstitümüz Yönetim Kurulunun 11.09.2014 tarih ve 2014/14 sayılı kararıyla oluşturulan jüri tarafından **oybirliği / oyçokluğu** ile Yüksek Lisans Tezi olarak **kabul** edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 02.10.2014

1) Tez Danışmanı : Yrd. Doç. Dr. Metin ZONTUL

2) Jüri Üyesi : Prof. Dr. Zafer ASLAN

3) Jüri Üyesi : Yrd. Doç. Dr.. Atakan ERDEM

Metin Zontul
Zafer Aslan
Atakan Erdem

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form imzalanacaktır. Aksi halde geçersizdir.

ÖNSÖZ

Günümüzde şirketler arasında artan rekabet, müşteriye ve müşteri yönetimini odak noktası haline getirmiştir. Artık şirketler kısa dönemli kar marjından daha çok uzun dönemli yatırımlar yapmaktadır. Bu tür uzun dönemli yatırımların başarısını ise şirketin, müşteri ile olan ilişkileri ve bu ilişkilerde ki müşteri tatmini belirler.

Müşterinin bu kadar odak noktası olduğu günümüzde müşteri yönetimi ile ilgili şirketlerin yatırım yapması kaçınılmaz bir durumdur. Geçmişte kullanılan müşteri yönetimi yöntemleri teknolojinin bugün ulaştığı noktada yetersiz kalabilmektedir. Bugün belki de müşteri yönetiminin en çok ihtiyaç duyduğu alan sosyal medyayı yönetebilmektir. İnsanlar sosyal medya aracılığıyla birçok hayalini ve sahip olmak istedikleri ürünleri paylaşmaktadır. Böyle bir müşteri havuzunu yönetebilmek her şirketin hayalidir.

Çalışma boyunca üzerinde duracağımız metin madenciliği ile sosyal medya kullanımı konusu müşteri yönetimi bazında oldukça önemli bir konudur. İnsanlar beğendiği ürünler ile ilgili fikirlerini arkadaş veya aile çevreleriyle sıkça paylaşmaktadır ve bu paylaşımlar olumlu olduğu gibi olumsuz da olmaktadır. Çalışmamızda sosyal medya üzerinde paylaşılan şikayetler üzerinden metin madenciliği kullanarak hangi firmanın hangi konularda ne kadar şikayet aldığına analizi yapılmıştır. Bu analiz sonuçları raporlanarak yüzdesel oranda uygulama ara yüzü ile kullanıcılara sunulmuştur.

Çalışmamızın birinci bölümünde tez konusu ile ilgili genel bir açıklama metni bulunmaktadır. İkinci bölümünde metin madenciliğinin ana dalı olan veri madenciliğinden bahsedilmektedir. Çalışmamızın üçüncü bölümünde ise metin madenciliği hakkında bilgiler ve işleyiş yöntemleri incelenmektedir. Sosyal medya üzerinden toplanan verilerin metin madenciliği kullanarak nasıl analiz edilip raporlandığını da tezimizin dördüncü bölümünde

incelenmektedir. Tezin son bölümünde ise yapılan çalışmaların sonuçları ve bu sonuçların neler kazandırdığından bahsedilmektedir.

TEŐEKKÜR

Çalıřmalarım süresince desteęini ve yardımını esirgemeyen, Deęerli hocam Yrd. Doç. Dr. Metin Zontul'a teőekkür ederim. Yaptıęım çalıřmanın bilimsel bir kimlięe bürünmüş olmasında büyük bir emeęi vardır.

Çalıřmalarım süresi boyunca ailemin bana karşı göstermiş olduęu anlayıř, sabır ve destek benim için çok anlamlıydı. Sonsuz teőekkürler.

İstanbul,2014

Gökhan Aravi

İÇİNDEKİLER

ÖNSÖZ.....	i
TEŞEKKÜR.....	iii
ŞEKİL LİSTESİ.....	vi
TABLO LİSTESİ.....	vii
1.GİRİŞ.....	1
2.VERİ MADENCİLİĞİ NEDİR?	6
2.1 VERİ MADENCİLİĞİNDE KULLANILAN MODELLER	7
2.2 VERİ MADENCİLİĞİ YÖNTEMLERİ.....	8
2.2.1 Sınıflandırma	8
2.2.2 Kümeleme	8
2.2.3 Birliktelik Kuralları	9
3.METİN MADENCİLİĞİ NEDİR?	10
3.1 METİN MADENCİLİĞİ OLUŞTURMA SÜREÇLERİ	10
3.1.1 Metin Toplama.....	10
3.1.2 Metin Önışleme	10
3.1.3 Veri Madenciliği	10
3.1.4 Raporlama ve Değerlendirme	11
3.2 METİN MADENCİLİĞİNİ OLUŞTURAN TEMEL ALANLAR VE FONKSİYONLARI	11
3.2.1 Metin Madenciliğini Oluşturan Temel Alanlar.....	11
3.2.2 Metin Madenciliğini Fonksiyonları.....	12
3.3 METİN MADENCİLİĞİ HAZIRLIK VE ÖNİŞLEME	13
3.3.1 Metin Madenciliği Hazırlık.....	13
3.3.2 Metin Madenciliği Önışleme	14
3.3.2.1 İşaretleme	15
3.3.2.2 Gövdeleme.....	16
3.3.2.3 Çok Kelimeli Özellikler	17
3.3.2.4 Kelime Anlamında Belirsizliğin Giderilmesi	17
3.3.2.5 Sözlük Oluşturma	18
3.3.2.6 Sözlük Türü Etiketleme	19
3.3.2.7 Öbek Tanıma	21
3.3.2.8 Sözdizimsel Analiz.....	21

4.SOSYAL MEDYA ÜZERİNE BİR METİN MADENCİLİĞİ UYGULAMASI	23
4.1 UYGULAMA AMACI	23
4.2 UYGULAMANIN SİSTEM YAPISI	23
4.2.1 WPF (Windows Presentation Foundation) Teknolojisi ve Faydaları ...	23
4.2.2 Entity Framework Teknolojisi ve Faydaları	25
4.3 VERİTABANI OLUŞTURMA AŞAMASI	26
4.3.1 SQL Tarafında Veritabanı Oluşturma	26
4.4 UYGULAMA VERİ TOPLAMA AŞAMALARI	30
4.4.1 Veri Toplama	30
4.4.2 Veri Temizleme ve Kategori Oluşturma	35
4.4.3 Gövdeleme	36
4.4.4 Sözlük Oluşturma	36
4.5 ALGORİTMA UYGULAMA	37
5. SONUÇ	45
KAYNAKÇA	47
WEB KAYNAKLARI	49
EKLER	50
ÖZET	58
ABSTRACT	59

ŞEKİL LİSTESİ

Şekil 3.1 Metin Madenciliği Kullanımı Alanları	11
Şekil 3.2 Metin Madenciliği Fonksiyonları.....	12
Şekil 4.1 Entity Framework Kullanma.....	27
Şekil 4.2 ADO.NET Entity Data Model Ekleme	28
Şekil 4.3 Entity Data Model Sihirbazı.....	29
Şekil 4.4 Entity Framework Model	30
Şekil 4.5 Nuget Uygulamasına Ulaşma.....	31
Şekil 4.6 Class Ekleme Adımı.....	32
Şekil 4.7 Uygulamanın Referans Listesi.....	32
Şekil 4.8 Başlık Yaz Metodu	33
Şekil 4.9 Detay Yaz Metodu	34
Şekil 4.10 ReplaceText Metodu.....	35
Şekil 4.11 Uygulama Sözlük Yapısı.....	37
Şekil 4.12 Sözlük Bilgisi Çekme.....	38
Şekil 4.13 Uygulama Şeması.....	38
Şekil 4.14 Metin Madenciliği Algoritmasının Akış Diyagramı.....	39
Şekil 4.15 Analiz Programı Raporlama Arayüzü - 1	42
Şekil 4.16 Analiz Programı Raporlama Arayüzü - 2	43
Şekil 4.17 Analiz Programı Raporlama Arayüzü - 3	44

TABLO LİSTESİ

Tablo 3.1 Sözcük Türü Kategorileri	20
Tablo 4.1 Kategoriler	35
Tablo 4.2 Kategoriler ve Sözcükler.....	37
Tablo 4.3 Yorum 1 Sonrası Analiz Sonucu.....	40
Tablo 4.4 Yorum 1 Sonrası Analiz Sonucu.....	40
Tablo 4.5 Yorum 1 Sonrası Analiz Sonucu.....	41

KISALTMALAR

WPF : Windows Presentation Foundation

EF : Entity Framework

1. GİRİŞ

Günümüzde gelişen teknolojilerle beraber arařtırmalar yapabilmek ve bu arařtırmalar için veriler elde edebilmek gittikçe kolaylařıyor. Kişisel bilgisayarların artması buna baęlı olarak internet kullanımının da artmasını saęlamıřtır. Gerek sosyal medya marifeti gerek ise forum siteleri üzerinden büyük çaplı veriler elde edilebilir duruma gelinmiřtir.

İnternetin bu veri kanalları ile bir insan grubunun nabzı ölçülebilir bir konu veya bir ürün hakkında ki düşünceleri ölçülebilir hatta beklentileri öngörülerek çalıřmalar yapılabilir. İnsan grubu olarak bahsettiğimiz bu topluluklar bir şehir halkını bir ülke halkını hatta dünya çapında bir insan topluluęunu oluřturuyor olabilir. Buna baęlı olarak elimizde bulunan veri inanılmaz büyük bir yığın haline gelecektir ve işlenmesi çok zor bir hal alacaktır. Bu tür veriler ancak belirli bir yordama göre işlendikten sonra anlamlı bir hal alacaklardır.

Veriler, yapılandırılmıř veriler ve yapılandırılmamıř veriler olarak iki grupta incelenebilir.

Yapılandırılmıř veriler (Structured Data) tablolar üzerinde satır ve sütunlar halinde düzenlenmiř verilerdir. Bunlar kâğıt üzerinde olabileceęi gibi, bilgisayarlarda istatistik paket programlarının satır ve sütunlarından oluřan matris şeklindeki yapılara da kayıt edilmiř olabilir. İstatistik yapılandırılmıř verilerle ilgilenmektedir. Bu tür verilerin sayısı çok olduęunda , verilerle veri madencilięi adı verilen bir dal ilgilenir (Oęuzlar, 2011, 1).

Veri madencilięi, büyük yığın verilerden anlamlı istatistikler, raporlar ve geleceęe yönelik tahminlerde bulanabilmemizi saęlamak için kullanılmaktadır. Bařka bir deyiřle Veri madencilięi, eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanıřlı bilginin çıkarılmasıdır. Bu da; kümeleme, veri özetleme, deęiřikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklařımları içerir [2].

Bir veri madenciliği sistemi, aşağıdaki temel bileşenlere sahiptir:

1. Veritabanı, veri ambarı ve diğer depolama teknikleri
2. Veritabanı ya da Veri Ambarı Sunucusu
3. Bilgi Tabanı
4. Veri Madenciliği Motoru
5. Örüntü Değerlendirme
6. Kullanıcı Arayüzü [2].

Yapılandırılmamış veriler (Unstructured Data) makale, köşe yazısı, email, web sayfası ve buna benzer metinlerden oluşabilir. Bu durumda Metin Madenciliği, istatistiğin yapılandırılmış veriler için yaptığı işi yapılandırılmamış veriler üzerinde yapmaktadır. Metin madenciliği veri madenciliğinin bir yan uygulama alanıdır.

Metin madenciliği, formatı olmayan metinlerin içinden bilgiler çıkarılması ve düzensiz verilerin formatlanması sürecidir. Metin madenciliği dört adımdan oluştuğu söylenebilir;

1. Metin madeni oluşturma
2. Metin işleme
3. Veri Analizi
4. Değerlendirme ve Raporlama

Metin madenciliğinden birçok alanda yararlanılmaktadır. Bu dallardan bir tanesi tıptır. Metin madenciliği tekniklerinin tıpta kullanımı son birkaç yılda büyük oranda artmıştır. Tıptaki verilerin genel olarak serbest metin formatında bulunması hasta ile ilgili önemli bilgilerin gözden kaçmasına, bilgiye erişimin zorlaşmasına sebep olmaktadır. Özellikle Elektronik Sağlık Kayıtlarının, Sağlık Bilgi Yönetiminin son yıllarda en önemli hedeflerinden biriyken, böyle bir sistemin başarısının, klinik dokümantasyonun serbest metin formatında yapılmasından dolayı sınırlanmış durumda olması bu tür sistemlere olan ihtiyacı ortaya çıkarmıştır. Yapılan klinik çalışmalar, araştırma raporları, hastane kayıtları, doktor notları, prosedürler ve faturalar tıptaki en

önemli veri kaynaklarıdır. Bu verilerin çoğu serbest metin formatında bulunmaktadır (Konchady , 2006).

Metin madenciliği, tıp alanında özellikle tıbbi araştırmalarda, semptomlarla hastalıklar ve ilaçlarla kimyasal maddeler arasında nedensel bağları bulmada, hasta kayıtlarının analiz edilmesinde, gen-gen ve protein-protein ilişkilerinin tanımlanmasında, tanı ve tedavileri geliştirmek, servis kalitesini ve faydayı arttırmak, maliyetleri kontrol etmek için kullanılmaktadır (Bhatt, 2004) [1].

Metin madenciliğinin kullanıldığı başka bir alan ise soru-cevap yöntemleridir. Doğal Dil İşleme disiplini altında yer alan biçimsel analiz yöntemine göre kullanıcıdan alınan soru metni işlenmektedir. Kullanıcıdan alınan soru metni için anahtar sözcükler belirlenip bu sözcüklerin her metin için ağırlığı belirlenmektedir. Bu ağırlık vektör uzay modelinde gösterilmektedir. Vektör uzay modeli bilgi çıkarımı, bilgi filtreleme, indeksleme gibi alanlarda kullanılan cebirsel bir modeldir. Doğal dil belgelerinin çok boyutlu uzayda özel bir anlamını simgelemektedir (Pılavcılar, 2007).

Metin madenciliği günümüzde popüler olan sosyal medya alanında da kullanılmaya başlanmıştır. Bunlara örnek olarak içinde Check-up geçen bir tweet attığınızda bir hastanenin sizi ulaşması veya bir araba modeli ile ilgili düşüncenizi paylaştığınız anda bir otomotiv şirketinin size ulaşip teklif vermesi gibi durumları gösterebiliriz. Bu kuruluşların sosyal medya kullanarak müşterilerine ulaşabilmesinin altında metin madenciliği yatmaktadır.

Sosyal medya üzerinden bazı analizlerde yapılabilmektedir. Bu çalışmada örnek olarak ele alınacak uygulamada insanların şirketlerin hizmetlerinden şikâyetçi olduğu durumları yazdığı bir sosyal medya ortamında yazılan yorumları ele alıp buradan belirlenmiş olan kategorilerden ne kadar şikâyet alındığı belirlenecektir .

Wu He, Shenghua Zha ve Ling Li'nin yaptıkları bir çalışmada pizza şirketleri arasında sosyal medya kullanılarak rekabet analizi çalışması

yapılmıştır. Ay bazında müşterilerin beğenileri, paylaşımları ve yorumları ele alınarak metin madenciliği marifeti ile analiz sonuçları ortaya koymuşlardır. (He, 2012).

Bir başka çalışma ise Kadriye ERGÜN tarafından yapılmış olan “Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi” tezi olmuştur. Bu tezde hepsiburada.com sitesi üzerinden çekilmiş yorumlar üzerinden metin madenciliği teknolojileri kullanılarak otomatik olarak puanlama sistemi yapılmıştır. Örnek uygulamamızda bizim kullanacağımız veriler sikayetvar.com sitesi üzerinden temin edilecektir (Ergün, 2012).

İnternet üzerinde ki yorumlar büyük önem arz etmektedir. Çünkü bir müşteri bu yorumları okuyarak ürün hakkında düşüncelerini büyük ölçüde şekillendirmektedir. Bununla birlikte tüketici davranışları;

- Kültürel etkiler,
- Bireysel etkiler,
- Grup etkileri

olmak üzere üç temel faktör üzerine kuruludur. Kültürel etkiler, davranış normları ve sosyalleşmeyi içerir. Bireysel etkilerde güdülenme, duygular, öğrenme ve hatırlama, tutum, algılama, rasyonel ve rasyonel olmayan düşünme, kişilik ve kişilik farklılıkları ve benlik özellikleri incelenir. Grup etkileri faktörü ise taklit etme ve öneri alma, aile, sosyal etkiler, etnik ve dinsel etkiler, sosyal sınıf, rol ve önderlerin etkisi gibi alt faktörlerden oluşmaktadır (Kayabaşı, 2010, 41-42).

Bunlara ek olarak

- Doküman sınıflandırma (Aşlıyan ve Günel, 2010; Gongde v.d., 2006),
- Benzer içerikleri belirleme (Karadağ ve Takçı, 2010),
- Web içerikleri sınıflandırma

alanlarında da metin madenciliği yöntemleri kullanılmaktadır.

Bu tez çalışmasında, metin madenciliğinin bir sosyal medya üzerindeki uygulaması yukarıda verilen dört adım takip edilerek gerçekleştirilmiştir. Örnek uygulama, diğer çalışmalardan farklı olarak bir şikâyet sitesindeki olumsuz yorumları inceleyip firmaların hangi alanlarda daha başarısız olduğunu görerek hizmetlerini hangi yönlerde geliştirmeleri ve hangi yönlere daha çok yatırım yapmaları gerektiğini görebilmeleri için tasarlanmış bir analiz sistemi çalışmasıdır.

2. VERİ MADENCİLİĞİ NEDİR?

Kavramsal olarak, “bir akıl yürütmenin, bir araştırmanın temeli olan ve olduğu gibi kabul edilen öge” olarak tanımlanan veri; ham, işlenmemiş, kullanılmak üzere olan olay veya durum olabilir (Bilgi Türleri, 1986).

“Veri”, yalnız başına anlamsızdır; ancak bir anlam yüklediğimiz takdirde kullanılabilir bir bilgi olacaktır. İstediğimiz amaç doğrultusunda bilgi oluşturmak için verinin işlenmesi gerekmektedir. Veriyi bilgiye çevirmeye “Veri analizi” denilmektedir. “Bilgi” ise bir soruya yanıt vermek için veriden çıkarılan sonuç olarak tanımlanabilir. Bu bakımdan veri ve bilgi arasında önemli bir iletişim ağı olduğu söylenebilir. Bilgi anlamlı olarak bazen yalın halde bazen de diğer bilgilerle birlikte çeşitli durumlarda istenilen sonuca ulaşmak için kullanılabilir (Alpaydın, 2000).

Veri Madenciliği büyük boyutlardaki verileri geleceğe yönelik tahmin yapmamıza imkan verecek bağlantıların ve modellerin meydana getirilmesidir. Bilgilerin günümüzde yoğun olarak kullanılan bilgisayar teknolojileri sayesinde sayısal ortamlarda saklamak ve bu bilgilere istenilen zamanda ulaşmak kolaylaşmıştır. Bu bilgilerden oluşan veri yığınlarından anlamlı sonuçlar çıkartmak ve geleceğe yönelik sağlıklı tahminler yapabilme ihtiyacı doğurmuştur. Bu noktada devreye veri madenciliği teknolojileri girmiştir.

Bilgisayar teknolojileri geliştikçe veri saklama maliyetleri düştü ve veritabanı sistemleri gelişti. Buna bağlı olarak bankacılık, tıp, otomotiv ve güvenlik gibi bir ülkede lokomotif olan sektörlerde veri madenciliği teknolojileri yoğun olarak kullanılmaya başlanmıştır.

Birinci aşamada ihtiyaçların belirlenmesi önemlidir. Stok analizlerinde satış yönetimi ve müşteri yönetimi alanlarda kullanılmak istenebilir.

İkinci aşama verilerin nasıl temin edileceğidir. Bu veriler şirket içerisinde kullanılan CRM, Stok yönetim programları gibi veri tutan uygulamalardan sağlanabilir.

Üçüncü aşamada ise veri sağlamak amacıyla hangi kayıtların ve alanların kullanılacağı belirlenmesidir. Gözlem sayısının çok fazla olması durumunda örnekleme yapılarak zaman ve maliyetlerden tasarruf edilir (Gürsoy, 2010, 4).

Kullanılacak verilere karar verildikten sonra veriler içerisindeki eksik veriler tamamlanır veya tamamen silinir. Buna ek olarak gürültülü verilerin incelenerek düzeltme yoluna gidilmelidir.

Modelleme aşamasında veriye uygun tahmin edici yöntemlere ve algoritmalara karar verilmelidir.

Sürecin bazı aşamalarında akışın çift yönlü olduğu dikkati çekebilir. Örneğin bilgi ihtiyacının belirlenmesi ve verilerin hangi kaynaklardan elde edileceğine karar verilmesi aşamalarında iki yönlü akış dikkati çekmektedir. Bu durum, sürecin bir aşamasında problem yaşanırsa, bir diğer adıma geri dönmek gerektiğini göstermektedir.

2.1 VERİ MADENCİLİĞİNDE KULLANILAN MODELLER

Veri madenciliği modelleri ikiye ayrılmaktadır.

- **Tanımlayıcı Modeller:** Amaç kadar vermeye yardımcı olacak desenlerin tanımlanmasını sağlamaktır. Geliri 3.000 ve 5.000 para birimi aralığında olan ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri 1.500 ve 3.000 para birimi aralığından düşük olan ailelerin satın alma örüntülerinin benzerlik gösterip göstermediğinin belirlenmesi tanımlayıcı modellere bir örnektir (Akpınar, 2000).
- **Tahmin Edici Modeller:** Tahmin edici modellerde (Predictive Models) sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak, sonuçları

bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesi amaçlanmaktadır. Örneğin bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere, sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır (Eker, 2005).

2.2 VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri Madenciliği modelleri temel olarak şu şekildedir:

- Sınıflandırma
- Kümeleme
- Birliktelik Kuralları

2.2.1 Sınıflandırma

Sınıflama veri madenciliği sıkça kullanılan bir yöntem olup, veri tabanlarındaki gizli örüntüleri ortaya çıkarmakta kullanılır. Verilerin sınıflandırılması için belirli bir süreç izlenir. Öncelikle var olan veri tabanının bir kısmı eğitim amacıyla kullanılarak sınıflandırma kurallarının oluşturulması sağlanır. Daha sonra bu kurallar yardımıyla yeni bir durum ortaya çıktığında nasıl karar verileceği belirlenir (Özkan, 2008, 45).

Örnek olarak genç yaştaki insanlar spor araba kullanmayı tercih ederken daha yaşlı insanlar daha çok büyük geniş ve lüks araba kullanmayı tercih etmektedir.

2.2.2 Kümeleme

Kümeleme analizleri sınıflandırma analizinden farklı olarak denetimsizdir. Alanların belirlenmesi ve birbirleriyle ilişkili verilerin alt kümeler oluşturması hedeflenmektedir.

Bir müşteri verisi ele alınırsa müşterilerin yaşları, ilgi alanları, cinsiyeti vb. birçok özellikler birbirleriyle karşılaştırılır ve ortak özelliklerinden alt kümeler oluşturulur. Oluşturulan bu kümeler belirli özelliklerde kitlelere hitap etme ve onlara uygun ürünler sunabilmek için kullanılabilir.

2.2.3 Birliktelik Kuralları

Birliktelik kuralları, birbiriyle ilişkili olan değişkenlerin ortaya çıkarılması ve aralarındaki bağlantının büyüklüğünün tespit edilmesine yöneliktir. Birliktelik kuralları belirli türlerdeki veri yapıları arasındaki ilişkileri tanımlamaya çalışan bir yöntemdir. Bağlantı analizleriyle cinsiyet ile eğitim durumu gibi çeşitli değişkenler arasında anlamlı ve kuvvetli bir bağlantı kurulabilir. Müşteri yaşı ve gelir seviyesi ile satın alma tutum ve davranışları arasında da bir bağlantı kurulabilir [5]. Bu tür bağlantılar ile müşterilerin alışkanlıkları, tercihleri belirlenmeye çalışılır. Bu bağlantılardan elde edilen bilgiler kullanılarak müşteri özel hizmetler ve uygun ürünler sunulabilir.

Bir müşterinin bankanın maaş müşterisi olduğunu düşünelim. Bu müşteri aynı zamanda maaş müşterisi olduğu bu bankanın kredi kartına sahip olmak istiyorsa aynı özelliklere ait olan diğer müşterilerde kredi kartı sahibi olmak isteyebilir. Bu bilgi üzerine bir pazarlama stratejisi kurulabilir.

3. METİN MADENCİLİĞİ NEDİR?

Metin Madenciliği; yararlı, ilginç ve daha önce bilinmeyen bilginin, bilgi işlem metotları ve teknikleri ile metin halindeki veriden elde edilmesi olarak tanımlanabilir.

3.1 METİN MADENCİLİĞİ OLUŞTURMA SÜREÇLERİ

3.1.1 Metin Toplama

Metin madenciliği uygulanacak metinlerin oluşturulma sürecidir. Metin toplama işlemleri genellikle internet üzerinden çevrimiçi bilgi havuzlarından yararlanılarak yapılmaktadır. Bu metinler herhangi bir meslek ile ilgili olabilir, güncel bir konu ile ilgili olabilir hatta bir köşe yazısı bile ele alınabilir.

3.1.2 Metin Önışleme

İşaretleme, gövdeleme sözlük oluşturma ve gereksiz kelimeleri ayıklama, yazım kurallarına uygunluğu tespit etme ve var olan hataları düzeltme gibi metin belgelerinin yapıtaşları olan kelimelerle ilgili işlemleri içeren süreçtir. Kimi çalışmalarda özellik seçme aşaması metin ön işlemenin bir aşaması olarak ele alınırken, bazı çalışmalarda ise ön işlemenin ardından gerçekleştirilen farklı bir adım olarak düşünülmektedir (Oğuzlar, 2011, 9).

3.1.3 Veri Madenciliği

Veri madenciliğinde, belirli bir formatta göre işlenmiş düzenlenmiş veriler üzerinde belirli algoritmalar kullanılır. Metin madenciliğinde bu durumun tersi olarak işlenmemiş veriler ile belirli desenler oluşturulmaya çalışılmaktadır. Bu işleme farklı bir deyişle Metin kategorizasyonu denebilir. Algoritmalar ise amaca yönelik oluşturulur ve kategorilendirilmiş metinlere uygulanır.

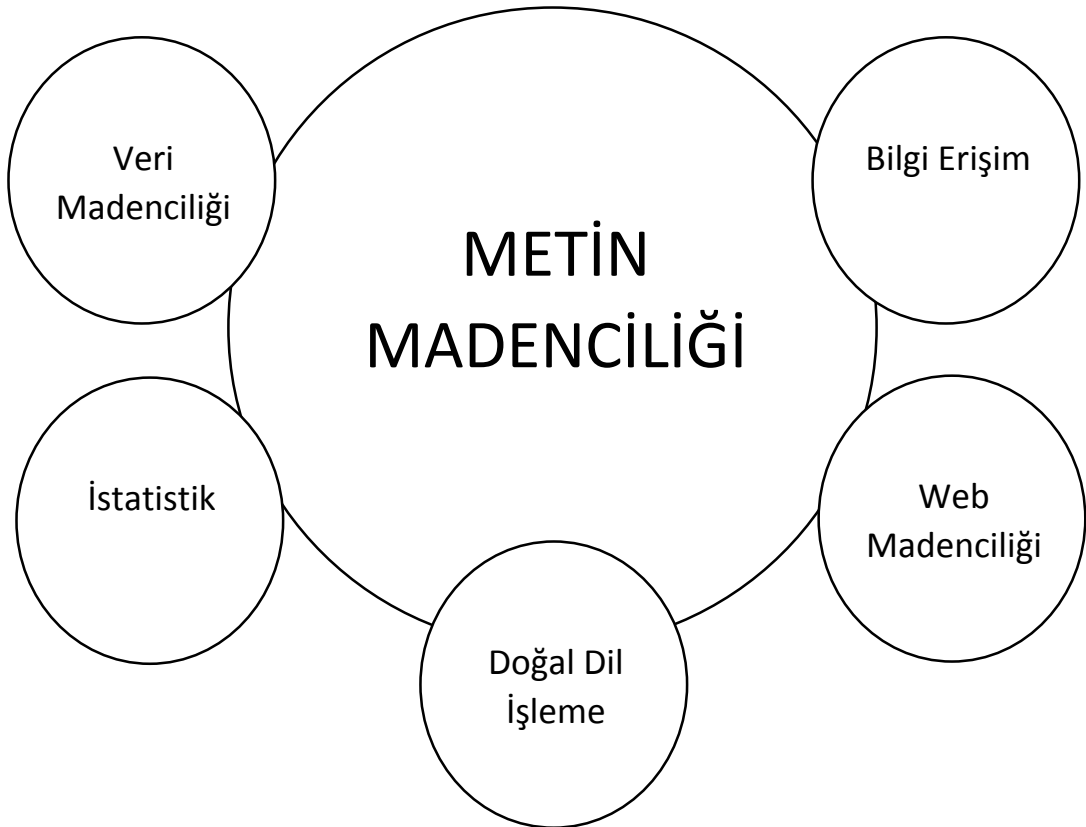
3.1.4 Raporlama ve Değerlendirme

Yapılan işlemler sonucunda elde edilen bilgiyi son kullanıcının anlayabileceği şekilde raporlama aşamasıdır.

3.2 METİN MADENCİLİĞİNİ OLUŞTURAN TEMEL ALANLAR VE FONKSİYONLARI

3.2.1 Metin Madenciliğini Oluşturan Temel Alanlar

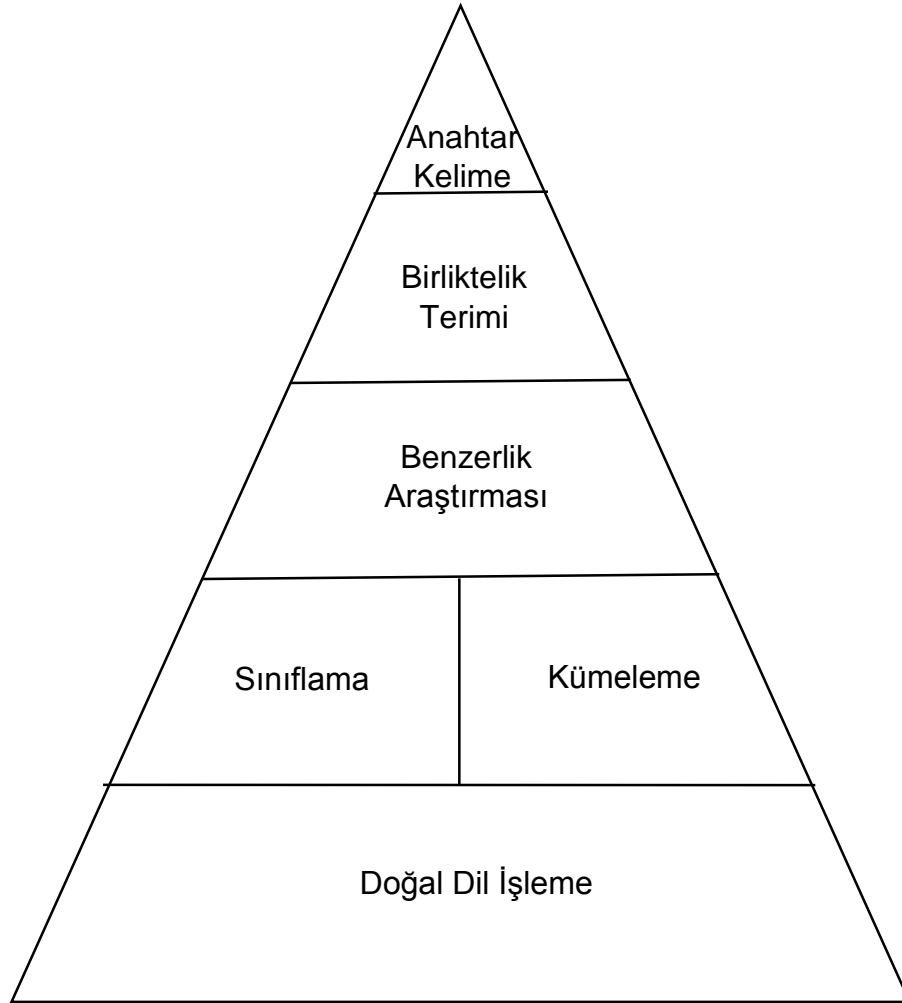
Metin madenciliği birçok alanda kullanılabilir olsa da genel olarak bilgi erişim ve istatistik, doğal dil işleme ve veri ve web madenciliği birleşimden oluşmaktadır. Şekil 3.1 'de temel alanlar gösterilmektedir.



Şekil 3.1 Metin Madenciliği Kullanımı Alanları (Chiwara, Al-Ayyoub, Hossain, Gupta, 2010)

3.2.2 Metin Madenciliğini Fonksiyonları

Metin Madenciliğinin fonksiyonları en basitinden en karmaşığına doğru Şekil 3.2’de gösterilmiştir.



Şekil 3.2 Metin Madenciliği Fonksiyonları (Oğuzlar, 2011)

Sınıflandırma işlemi nesnelerin daha önceden bilinen sınıflara ya da kategorilere dâhil edilmesidir. Birliktelik analizi ise sıklıkla birlikte yer alan ya da gelişen sözcük ya da kavramların belirlenmesi ve böylece doküman içeriğinin ya da doküman kümelerinin anlaşılmasını amaçlamaktadır. Bilgi çıkarım teknikleri ile dokümanların içerisindeki yararlı veri ya da ifadeler

bulunmaya çalışılmaktadır. Kümeleme analizi, doküman kümelerinin temelini oluşturan yapıların keşfedilmesi amacıyla uygulanmaktadır.

3.3 METİN MADENCİLİĞİ HAZIRLIK VE ÖNİŞLEME

3.3.1 Metin Madenciliği Hazırlık

Metin madenciliği, doküman koleksiyonları ile başlamaktadır. Bunlara en basit şekliyle “derlem” (corpus) denmektedir. Bu koleksiyonlar geleneksel veri tabanı ile kıyaslandığında, doküman koleksiyonu yapısal olmayan ham verilerden oluşmaktadır. Bu verile özel bir bilgisayar dilinde olabileceği gibi, doğal dilde de yazılmış olabilir. Derlem içerisinde yer alan dokümanlar paragrafları, paragraflar cümleleri ve cümleler ise kelimeleri içermektedir. Bunun yanında derlemenin yapısı statik ve dinamik olabilmektedir. Derlemlerin eğer başlangıçtaki durumları değişmeden kalıyorsa statik yapıda oldukları söylenebilir. Buna karşın zaman sürecinde yeni dokümanlar ekleniyor veya dokümanlar güncelleniyorsa derlemin dinamik olduğu söylemek mümkün olacaktır (Feldman ve Sanger, 2007, 2). Bir dokümanın anlamı, diğer bir deyişle gerçekte ne söylediği ise kelimelere bağlıdır (Miller, 2005, 105).

Bazı durumlarda ise veri toplam süresine ihtiyaç duyulabilir. Yapılacak bir anket, forumlar ve web siteleri aracılığıyla verilerin toplanma süreçleri uzun sürebilir. Farklı bazı uygulamalarda ise, girdi veri akışına uzun bir süre müdahale edilerek bir süreç içinde veri toplanabilir (weiss, Indurkha, Zhang ve Damerou, 2010, 16.). Giriş bölümünde bahsedildiği gibi internet marifeti ile veriye ulaşım kanal sayısı artmış ve eskiye göre çok daha kolaylaşmıştır. Toplanılan veri ise belirli bir formatı getirebilmemiz için bir ön işlemine tabi tutulmalıdır.

3.3.2 Metin Madenciliği Önışleme

Etkili ve gelişmiş bir metin madenciliği işleme yapılabilmesi için veriler ön işleme metotlarına tâbi tutulmalıdır. Çok sayıda ve farklı metin madenciliği ön işleme tekniği mevcuttur (Buss, 2011).

Önışleme tekniklerinin görev odaklılığı ya da türetildikleri biçimsel çerçevelere göre kategorilere ayrılması, belirli bir metin madenciliği uygulaması için herhangi bir kategorideki tekniklerin “bir arada kullanılması ve eşleştirilmesi” işlemlerinin yasak olduğu anlamına gelmektedir. Metin madenciliği ön işleme faaliyetlerindeki algoritmaların çoğu belirli görevlere özgü değildir ve problemlerin çoğu birbirinden oldukça farklı olan birçok algoritma ile çözülebilmektedir (Feldman ve Sanger, 2007, 57).

Ön işleme tekniklerinden her biri kısmen yapılandırılmış bir dokümanla işe başlamakta ve mevcut özellikler rötuş edilerek ve/veya yeni özellikler eklenerek yapı zenginleştirilmektedir. Sonunda en gelişmiş olan ve anlamı en iyi temsil eden özellikler metin madenciliğinde kullanılırken, geriye kalanlar kapsam dışı bırakılmaktadır. Girdiyi oluşturan görünüm ve çıktıyı oluşturan özelliklerin karakteristiği, ön işleme teknikleri arasındaki temel farklılığı oluşturmaktadır (Feldman ve Sanger, 2007, 58).

Metin Madenciliği önışleme yöntemleri:

- İşaretleme
- Gövdeleme
 - Joker Yöntemi
 - Köke Kadar Gövdeleme
- Çok Kelimeli Özellikler
- Kelime Anlamında Belirsizliğin Giderilmesi
- Niteliklerin Sıralandırılmasıyla Özellik Seçimi
- Sözlük Oluşturma

- Sözcük Türü Etiketleme
- Öbek Tanıma
- Sözdizimsel Analiz

3.3.2.1 İşaretleme

Metin ile ilgili çalışmada atılacak ilk adım, karakter akışını kelimelere, ya da daha net bir deyişle işaretlere bölmektir (Weiss, Indurkha, Zhang ve Damerau, 2010,16). Üst düzey işleme durumlarında işaretleme yöntemi kullanıma zorunluluğu doğurmaktadır. Metin madenciliği sistemlerinde işaretleme düzeyleri genelde metnin cümle ve kelimelere ayrılmasıyla ilgilidir (Feldman ve Sanger, 2007, 58).

Karakter akışının işaretlere bölünmesi, dil yapısını bilen biri için sıradan bir görevdir. Ancak dili algılama görevi verilen bir bilgisayar programı için bu görev daha zor olmaktadır. Bunun sebebi belirli karakterlerin kullanımına bağlı olarak, bazı durumlarda işaretin sınırlayıcı olması, bazı durumlarda ise olmamasıdır. Boşluk, sekme ve satırbaşı karakterlerinin hep sınırlayıcı olduğu, işaret olarak sayılmadığı varsayılmaktadır. Bunlara sıkça toplu olarak beyaz boşluklar adı verilir.”()”,”< >”,”!” ve “?”karakterleri her zaman sınırlayıcı olup, aynı zamanda işaret olabilmektedir.”.”,”,”,”:” ve “-“ karakterleri, ortamlarına bağlı olarak sınırlayıcı olabilir veya olmayabilir. Numaralar arasındaki nokta, virgül ya da iki nokta üst üste normalde sınırlayıcı değil, numarasının bir parçası sayılmaktadır. Her türlü başka virgül ya da iki nokta üst üste sınırlayıcı olup işaret olabilmektedir. Nokta, kısaltmanın bir parçası olabilir, (örneğin “Dr.”). Ancak bunlardan bazıları aslında cümle sonudur. İşaretleme (tokenizasyon) amaçları için belki de en iyi çözüm, belirsiz her noktayı kelime sınırlayıcı ve aynı zamanda işaret olarak ele almaktır (Oğuzlar, 2011, 32).

3.3.2.2 Gövdeleme

Veri kümesi işaretleme işlemi uygulandıktan sonraki adım, işaretlerden her birini standart hale getirilmektir. Bu adımın uygulanması duruma göre değişiklik gösterebilir. Çünkü bazı durumlarda gövdeleme bir kazançtan çok kayıp yaratabilmektedir. Gövdeleme iki ana başlıktan oluşmaktadır:

- **Joker Yöntemi:** Türkçe için en uygun yöntemdir. Bunun sebebi Türkçe gibi eklemeli dillerde bir gövdenin sonuna birçok farklı ek alarak farklı biçimlerde karşımıza çıkabilmesidir. Örneğin “araba” kelimesi ile “arabada”, “arabayı”, “arabada”, ve “arabanın” kelimeleri eğer ayırıştırıcı olmasa ayrı ayrı kelimeler olarak görüleceklerdi. Bunun sonucu olarak hem oluşturulan sözlük boyutu çok artacak hem de sınıflandırma başarısı düşecekti. Joker kelime, aynı söz dizimi ile başlayan ve çeşitli ekler almış ancak yakın anlamda olan sözcükleri tek bir gösterimle grup altında toplayan kelimelerdir. Joker kelime gövdeleme yöntemine benzerdir. Gövdelemede çekim ve yapım eklerinden ayırıştırılan kelimeler, ortak bir köke indirgenir. Ancak burada köke indirgeme şartı yoktur. Kökün yanında ek de kalabilir. Joker kelimeler kategoriyi belirlememize yardımcı anahtar kelimelerden veya sık kullanılan kelimelerden seçilir. Joker yöntemi kelimelerin ilişkili terimlerinin anlamlarını kapsamı açısından değiştirilmesidir. Örneğin jokerli bir kelime olarak “depren*” ile (Joker olduğunu * işaretinden anlıyoruz), “depren” kelimesinden sonra nasıl bir ek gelirse gelsin depren kelimesi vurgulanmış olacaktır.
- **Köke Kadar Gövdeleme:** Köke kadar gövdeleme yöntemi joker yöntemine göre daha katı kurallardan oluşan bir yöntemdir. Bu

yöntemde herhangi bir ek ile ilgilenmeden kök dizine ulaşması hedeflenmektedir.

Katı gövdelemenin sonucu, metin içindeki tip sayısının şiddetli bir şekilde azaltılması ve böylece dağılım istatistiğinin daha güvenilir yapılmasıdır. Yeterli kaynaklar mevcut olduğu sürece, şüphe durumunda hem gövdelemeli hem de gövdelemesiz olarak deneme yapmak yararlı olacaktır (Oğuzlar, 2011, 37).

3.3.2.3 Çok Kelimeli Özellikler

Çok kelimeli Özelliklerin değerinin ölçümü genel olarak, olası çok kelimeli özelliklerdeki kelimeler arasındaki korelasyonlar ele alınarak yapılmaktadır. Bu amaçla, ortak bilgi ya da benzerlik oranı bazında çeşitli ölçümler kullanılabilir.

Diğer farklı uygulamalar çok kelimeli özellikler oluşturulmadan önce, durdurucu kelimelerin kapsam dışına çıkarılıp çıkarılmayacağına bağlı olarak oluşturulabilir. Genellikle, doküman koleksiyonunda çok kelimeli özelliklere sıkça rastlanmamakta, fakat rastlanıldığında bu tür özelliklerin tahmin etme kapasitesi çoğu zaman yüksek olmaktadır. Çok kelimeli ifadeleri kullanmanın dezavantajı, metin ön işleme karmaşıklığını bir üst seviyeye çıkarmalarıdır. Bazı araştırmacılara göre, çok kelimeli özellikleri oluşturmaya yönelik bir ön işleme adımı gereksiz olup kelimelere bir araya getirmek öğrenim yöntemlerinin yapması gereken görevdir. Ancak öğrenim yöntemi bunu yapamamakta ise, ek çabanın sarf edilmesi yararlı olabilmektedir, çünkü çok kelimeli ifadeler çoğu zaman yüksek tahmin kapasitesine sahip olmakta ve sonuçların yorumlanabilirliğini arttırmaktadır (Weiss, Indurkha, Zhang ve Damerau, 2010, 34).

3.3.2.4 Kelime Anlamında Belirsizliğin Giderilmesi

Sözlüklerin, asli olarak işlevi kelimelerin anlamlarının kaydını tutmaktır. Fakat bu sözlükler dijital dünyada bir uygulamanın belirsizliğinin gidermesi

için kullanılacak şekilde oluşturulmamıştır. Kelimelerin anlamları ve aralarındaki ilişkilere odaklanmış büyük ve uzun vadeli proje olan Word net, bu boşluğu doldurmayı amaçlamıştır [3]. Türkçe Word Net'in oluşturulması projesi Sabancı Üniversitesi tarafından yürütülmektedir [6].

Fakat Word net başarılı bir proje olmuş olsa da yazılardaki anlam belirsizliklerini çözümlen bir algoritma ortaya koyamamıştır. Metin madenciliği projesinde gerek olmadıkça bu adımın uygulanmaması iyi bir seçenek olacaktır.

3.3.2.5 Sözlük Oluşturma

Bir dilin veya dillerin kelime haznesini (sözvarlığını), söyleyiş ve yazılış şekilleriyle veren, kelimenin kökünü esas alarak, bunların başka unsurlarla kurdukları sözleri ve anlamlarını, değişik kullanışlarını gösteren eserlere sözlük denir. Bir başka kaynağa göre sözlük, bir dilin veya dillerin kelime haznesini (sözvarlığını), söyleyiş ve yazılış şekilleriyle veren, sözcüğün kökünü esas alarak, bunların başka unsurlarla kurdukları sözleri ve anlamlarını, değişik kullanışlarını gösteren yazılı eserdir [7].

Öncelikli olarak sözlüklerin boyutlarını düşürmek için durdurucu kelimelerin (stopwords) bir listesi hazırlanır ve bu kelimeler sözlükten çıkartılır. Durdurucu kelimelere örnek olarak zamirler gösterilebilir.

Sözlük boyutunu küçültme teknikleri şu şekildedir;

- Lokal sözlük
- Durdurucu kelimeler
- Sıkça karşılaşılan kelimeler
- Özellik seçimi
- İşaret indirgeme: gövdeleme, eşanlımlı sözcükler (Weiss, Indurkha , Zhang ve Damerau, 2010, 27)

Kelime sayısını düşürme performansı arttıracığı gibi tahmin doğruluğunu da arttıracaktır. Bazı öğrenim yöntemlerinde büyük sözlük tiplerinden daha çok küçük sözlük tipleri daha etkili olabilmektedir.

Sözlük oluşturulurken bir kelimenin her varyasyonu bulundurmamak gereksiz olabilir. Bir kelimenin hem tekil halini hem de çoğul eki almış halini bulundurmamak anlamlı olmayabilir. Gövdeleme işleminde köke yaklaştıkça kelime anlamında değişimler olduğu durumlarla karşılaşılabilir. Yine de gövdeleme işlemi sözlük boyutlarında büyük bir küçülme sağlamaktadır.

3.3.2.6 Sözlük Türü Etiketleme

POS kullanıldığı cümlelerin içinde üstlendiği role göre kategorilere ayırmaktır. POS, kelimelerin içeriğine bağlı bilgi sağlar. POS etiketi kelimeleri genellikle standart kategorileri kullanmaktadır. Örneğin, Brown Corpus etiket kümesinde en az 87 adet temel etiket bulunmaktadır. Büyük çaplı metinlerde kullanılması önerilmektedir.

Tüm dillerde sözcükler türlere ayrılmıştır. Türkçeyi ele alacak olursak bu kategorilere örnek olarak isim, fiil, sıfat, zarf zamir ve bağlaç gösterilebilir. Sözcükler en azından bu kategoriler altında toplanabilmektedir. 8 Temmuz 1889 tarihinde kurulan Wall Street Journal gazetesinin metinlerden oluşturulmuş Penn Tree Bank 36 kategoriden meydana gelmiştir. Bu kategoriler aşağıda Tablo 3.1'de gösterilmiştir.

Tablo 3.1 Sözcük Türü Kategorileri [9]

No.	Etiket	Tanım
1.	CC	Düzenleme bağlacı
2.	CD	Nicelik sayısı
3.	DT	Niteleyici sözlük
4.	EX	Mevcudiyeti ifade etmek için kullanılan
5.	FW	Yabancı sözcük
6.	IN	İlgeç ya da yantümce bağlacı
7.	JJ	Sıfat
8.	JJR	Sıfat, üstünlük dereceli (komparatif)
9.	JJS	Sıfat, en üstünlük dereceli (süperlatif)
10.	LS	Liste kalemi işareti
11.	MD	Moda (Kip)
12.	NN	Tekil ya da toplu isim
13.	NNS	Çoğul isim
14.	NNP	Özel isimler, Tekil
15.	NNPS	Özel isimler, Çoğul
16.	PDT	Ön belirteç
17.	POS	İyilik durumunu gösteren son ek
18.	PRP	Kişi Zamiri
19.	PRP\$	İyelik Zamiri
20.	RB	Zarf
21.	RBR	Karşılaştırmalı Zarf
22.	RBS	Üstünlük Zarfları
23.	RP	Edat
24.	SYM	Simge
25.	TO	e, a, ye, ya
26.	UH	Ünlem
27.	VB	Mastar halinde fiil
28.	VBD	Geçmiş zamandaki fiil

29.	VBG	Bağfiil ya da durum ortacı biçimindeki fiil
30.	VBN	Geçmiş zaman ortacı biçimindeki fiil
31.	VBP	Üçüncü kişi olmayan tekil şimdiki zaman fiili
32.	VBZ	Üçüncü Tekil şahıs şimdiki zaman fiil
33.	WDT	Wh-belirteci
34.	WP	Wp-zamiri
35.	WP\$	WH iyelik zamiri
36.	WRB	Wh-zarfı

3.3.2.7 Öbek Tanıma

Bu adım kelimelerin öbek kısımlarına göre kümelenmesidir. Öbek tanıma cümleleri belli bir oranda analiz yapmak için kullanılması gereken yararlı bir adımdır.

3.3.2.8 Sözdizimsel Analiz

Sözdizimsel analiz, cümle içinde geçen her kelimenin diğer kelimelere olan bağlantısı ve cümle içinde ki rolünü bulabilmektedir. Her dilin kendine göre birçok farklı Sözdizimsel analizi türü bulunmaktadır.

Birçok geleneksel dilbilgisinde öbekler arasında bir ayırım yapılır. Örnek olarak ilgeç, sıfat, fiil vb. gibi öbek tipleri gösterilebilir. Öbekler dilin kurallarına göre sıfır veya kelimelerden oluşabilir. Büyük çaplı metinlerde kullanılması önerilmektedir.

Bu noktaya kadar bahsedilen ön işleme yöntemleri, metin madenciliği uygulanırken yararlı olabilecek metotların belirlenmesi ve kelime tiplerinin tespitinde sağlanacak faydalar ile ilgilidir.

3.3.3 Çalışma Tablosu Oluşturulması

Ön işleme metotlarından sonra oluşturulan koleksiyonlar çalışma tablosuna çevrilmek üzere belirli algoritmalarından geçtikten sonra, çalışma tablosuna her bir satır bir doküman ve her sütun bir özelliğe denk gelecek şekilde yerleştirilir.

Bu noktadan sonra yapılması gereken işlem belirlenen kelimelerin doküman içerisinde bulunup bulunmadığını incelemektir. Her kelimenin karşısında bulunup bulunmadığına bağlı olarak bir veya sıfır şeklinde belirtilmelidir. Bu yöntemden farklı olarak kelimelerin karşısına kelimenin görülme sayısı da yazılabilir. Üçüncü bir yöntem olarak ta kelimelerin görülme sayısı için bir eşik konulabilir. Bu kelimenin bir değer ifade etmesi için bu eşiği aşması beklenebilir. Bu eşiği aşmayan kelimeler ise sıfır kabul edilir.

4.SOSYAL MEDYA ÜZERİNE BİR METİN MADENCİLİĞİ UYGULAMASI

4.1 UYGULAMA AMACI

Bu uygulamanın amacı gıda hizmeti veren firmaların aldığı şikâyetleri belirli kategorilere bölerek hangi konuda daha çok şikâyet aldığını metin madenciliği yöntemlerini kullanarak belirlemek ve raporlamaktır.

4.2 UYGULAMANIN SİSTEM YAPISI

Uygulama Visual Studio 2013 kullanılarak C# yazılım dili altyapısı ile geliştirilmiştir. Projemiz veri ambarı ve analiz uygulaması adında iki projeden oluşmaktadır. Veri ambarı uygulaması veri toplayıp veritabanı üzerine kaydetmektedir. Analiz uygulaması ise veritabanı üzerinde bulunan verileri sözlük kullanarak analiz ederek bir arayüz üzerinden kullanıcıya yorumlayıp raporlamaktadır. Arayüz WPF teknolojisi kullanılarak hazırlanmıştır. Bu iki projede de veritabanı ilişkileri ve bağlantıları Entity Framework altyapısı kullanılarak gerçekleştirilmiştir. Kullanılan teknolojiler kısaca aşağıda bahsedilmiştir.

4.2.1 WPF (Windows Preentation Foundation) Teknolojisi ve Faydaları

WPF, 2003 yılında lanse edilmiş uygulama arayüzü geliştirme ortamıdır. Bu ortamın geliştiricilere sağlamış olduğu modern tasarım ekranları kullanıcıyla daha etkileşimli, veri gösteriminde daha esnek ve windows form üzerinde geliştirilmesi zor ekranların daha hızlı ve kolay şekilde ortaya çıkmasını sağlamaktadır. WPF, mimarisi gereği sistem kaynaklarını çok verimli kullanabilmektedir. Platformunun bu özelliği geliştiricinin en zor ve karmaşık uygulamalarda bile performans konusunda başarılı sonuçlar almasını sağlamaktadır (Yağmur, 2009). WPF, tasarım altyapısı XAML işaretleme diline dayandığı için bazı durumlarda değişiklikler ve eklemeler yapmak zor olabilir. Bu durum için Microsoft'un

geliştirmiş olduğu Expression Blend uygulamasını kullanarak çok rahat bir şekilde formlar üzerinde değişiklik ve eklemeler yapılabilmektedir.

WPF Özellikleri

1. İlk olarak 2003'te Avalon ismi ile duyurulmuştur ve Framework 3.0 ile geliştiricilere sunulmuştur.
2. WPF vektörel bir yapıya sahiptir.
3. WPF mimarisi görüntü işleme konusundaki başarısının arkasında yatan mimari directx kütüphanesidir. Bu kütüphane sayesinde hızlı görüntü işleme mümkün olmuştur. Windows form mimarisinde kullanılan GDI+ kütüphanesinden vazgeçilmesinde ki sebep directx kadar ekran kartı özelliklerini verimli kullanamamasıdır.
4. WPF görsel tasarım için XAML işaretleme dilini kullandığı için dinamik ekranlar tasarlanabilmektedir.
5. WPF tasarlanırken nesne tabanlı modelleme kurallarına uygun olarak geliştirildiği için Windows form tarafında yapılan birçok hata tekrar edilmemiştir.
6. WPF ile tarayıcı üzerinde çalışan uygulamalar geliştirilebilmektedir. XBAP/WPF Browser Application teknolojisi olarak bilinen bu özellik Java dilinde kullanılan java applet teknolojisine karşılık gelmektedir.
7. Silverlight eklentisi sayesinde flash uygulamalarına benzer uygulamalar geliştirilebilmektedir.
8. WPF ve Silverlight ile birlikte Expression adı verilen bir ürün paketi

geliştirilmiştir. Bu paket her iki platform için geliştirme araçları sağlamaktadır.

9. WPF teknolojisi kullanılarak geliştirilmiş uygulamalar sistemin merkezi birimleri olan işlemci ve bellek yerine ekran kartının işlemci ve belleğini kullanmaktadır. Bu duruma bağlı olarak uygulamalar sistemin merkezi birimini meşgul etmemektedir.

10. WPF, Windows Vista, Windows 7 ve Windows 8 gibi Microsoft'un geliştirmiş olduğu yeni nesil işletim sistemleri ile sorunsuz çalışmaktadır. Windows XP üzerinde WPF uygulamalarının çalışmasında sıkıntılar olabilmektedir. Windows XP'den önceki sistemlerde ise WPF uygulamaları çalışmamaktadır.

11. WPF, içerisinde gelen 3D kütüphanesi sayesinde doğrudan OpenGL ve directx kullanımı gerektirmemektedir.

12. WPF, herhangi bir çözünürlükte ve aygıtta bağımsız çalışabilme özelliğine sahiptir.

13. WPF, windows form kütüphaneleri ile uyumludur [8].

4.2.2 Entity Framework Teknolojisi ve Faydaları

Birbiriyle ilişkisi olmayan nesnelere iletişime sokmak, bu iletişimden doğan ilişkileri yönetmek, diyagram kullanarak bu nesnelere oluşturmak veya nesnelere kullanarak diyagram oluşturmak için Microsoft tarafından geliştirilmiş ve .net framework içerisine entegre edilmiştir [4].

Entity Framework üç türlü kullanım şekli bulunmaktadır;

1. Code First
2. Database First
3. Model First (Uzunköprü, 2014,1163)

Code First: Bu yöntemde model ve sınıfları yazılımcı oluşturur. Var olan database üzerinde revizyona gidilir.

Database First: Bu yöntemde varolan database dosyası proje üzerinde model dosyasına bağlanır. Sınıflar Entity Framework tarafından otomatik oluşturulur.

Model First: Bu yöntemde model dosyası projede oluşturulur. Bu model üzerinde sınıflar oluşturulduktan sonra Entity Framework database tarafında tabloları otomatik oluşturur.

4.3 VERİTABANI OLUŞTURMA AŞAMASI

4.3.1 SQL Tarafında Veritabanı Oluşturma

Bu aşamada MSSQL üzerinde sözlüğü oluşturan kelimeleri, bu kelimeleri sınıflandıracak kategorileri ve işlenecek verilerin tutulacağı veritabanı oluşturulur. Bu tezdeki veritabanı 4 tablo ve 4 procedure'den oluşmaktadır. Bu tablo ve prosedürler aşağıdaki gibidir;

Tablo

- ScreenName
- Status
- Texts
- Words

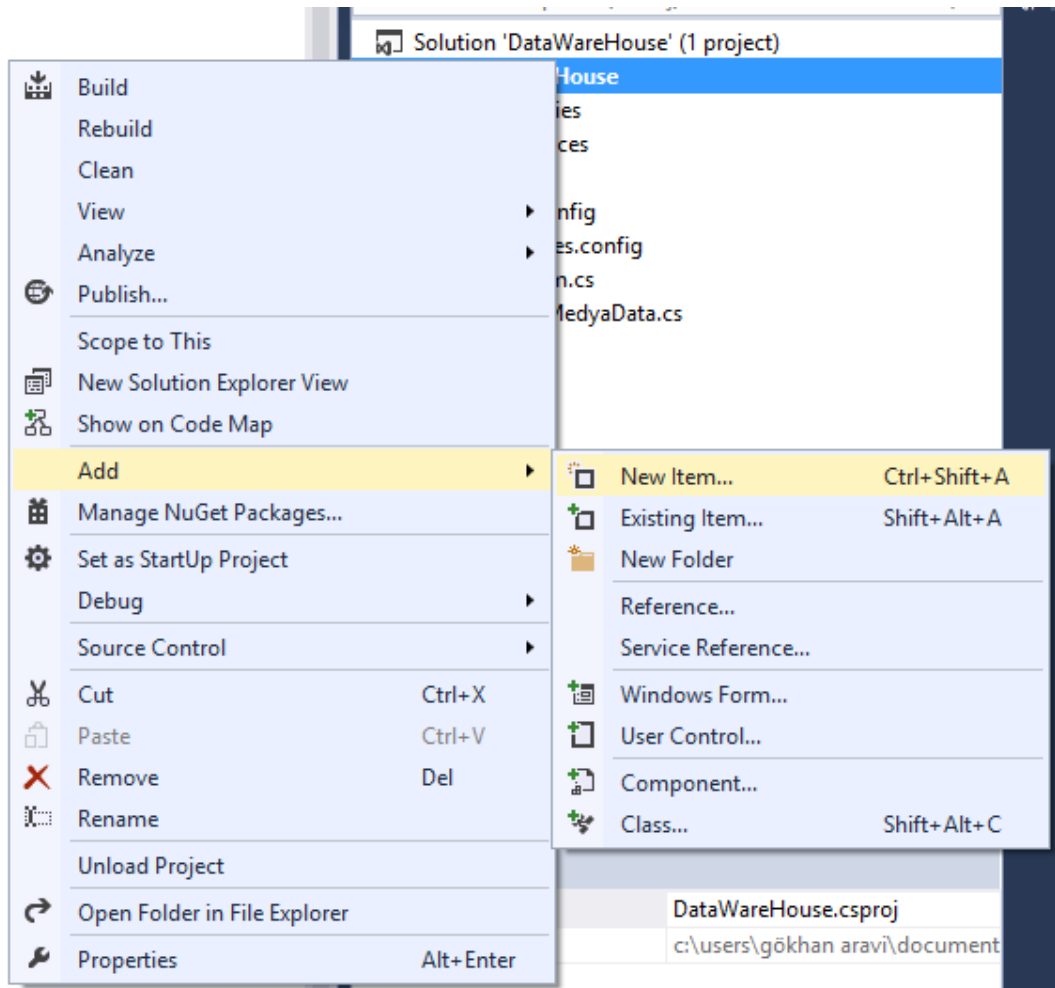
Procedure

- Word_List
- ScreenName_List
- Comment_List
- Comment_Add

Bu veritabanı Entity Framework kullanarak tezin uygulama bölümüne entegre edilmiştir.

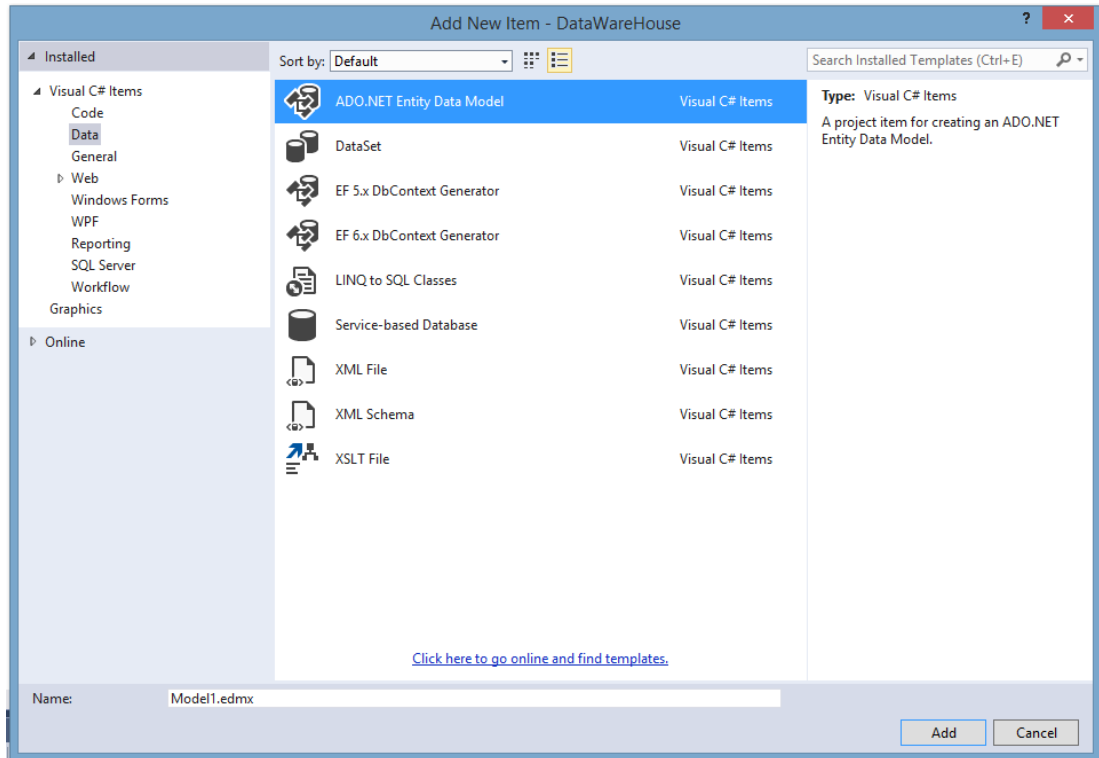
4.3.2 Entity Framework Kullanımı

Öncelikle projeye sağ tıklayıp edmx model dosyasını şekil 4.1'de ki gibi eklenir.



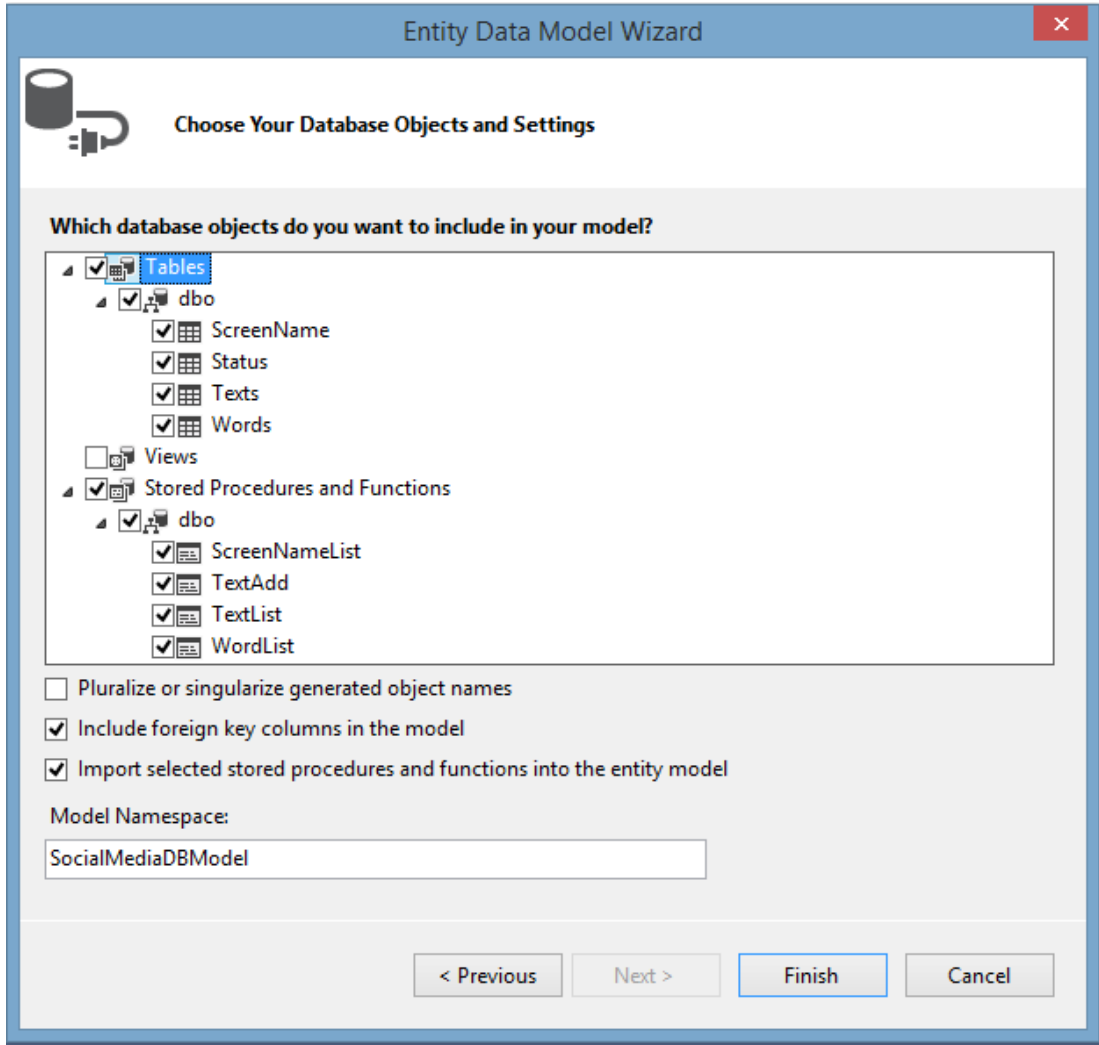
Şekil 4.1 Entity Framework Kullanma

Şekil 4.2'de görüldüğü gibi ulaşılan ekranda Data sekmesini seçtikten sonra ADO.NET Entity Data Model aracı uygulamaya eklenir.



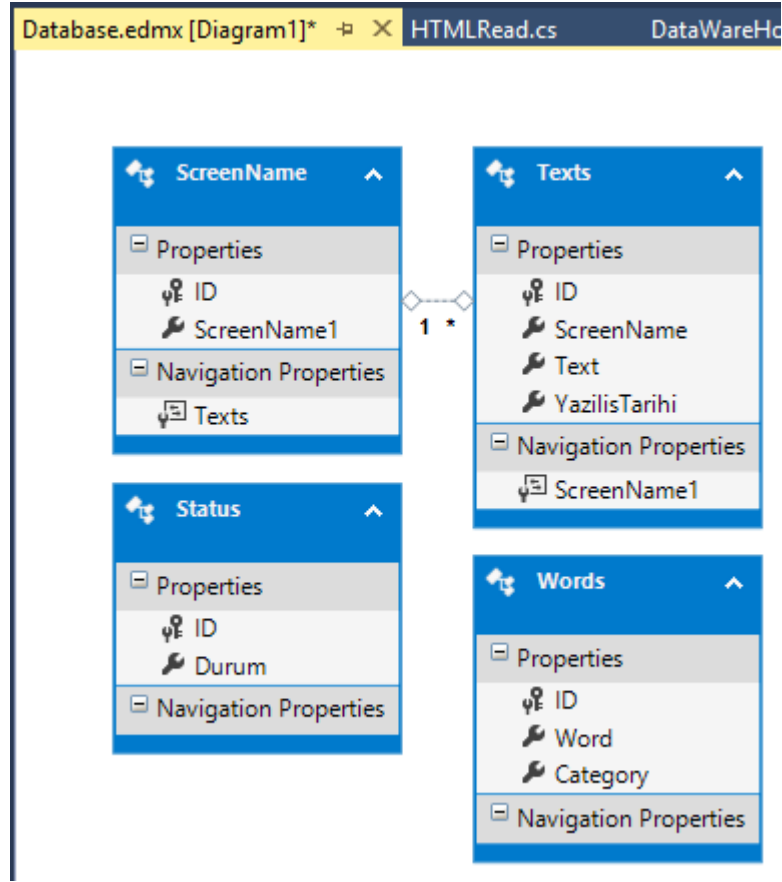
Şekil 4.2 ADO.NET Entity Data Model Ekleme

Add butonuna bastıktan sonra Visual Studio 2013 üzerinde standart ConnectionString oluşturma ekranında veritabanı bağlantısını gerçekleştirilir. Daha sonra Şekil 4.3'te görülen ekrandan veritabanı üzerindeki tablolar, procedureler ve viewler listelenir. Bu projede kullanılacak objeleri seçip devam tuşuna basılır.



Şekil 4.3 Entity Data Model Sihirbazı

Finish butonuna bastıktan sonra Şekil 4.4 ekranına ulaşılır ve uygulama üzerinde oluşan objeler görülür.

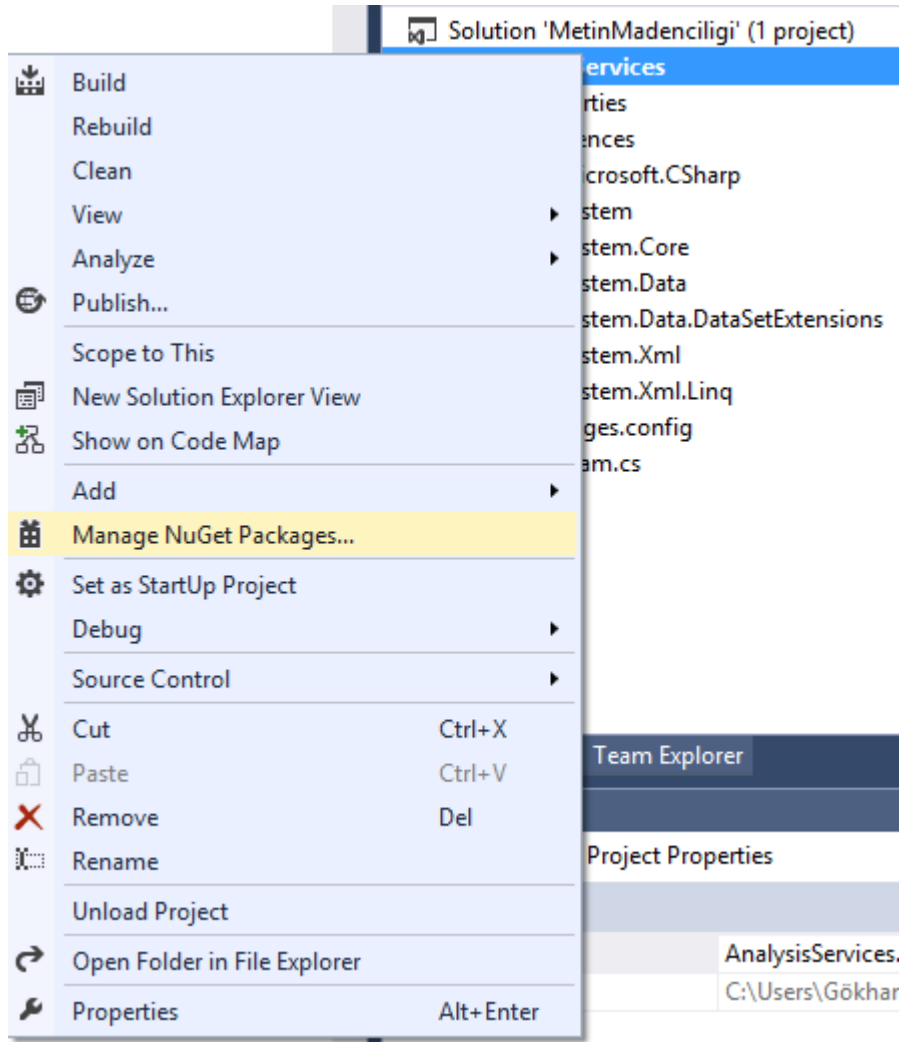


Şekil 4.4 Entity Framework Model

4.4 UYGULAMA VERİ TOPLAMA AŞAMALARI

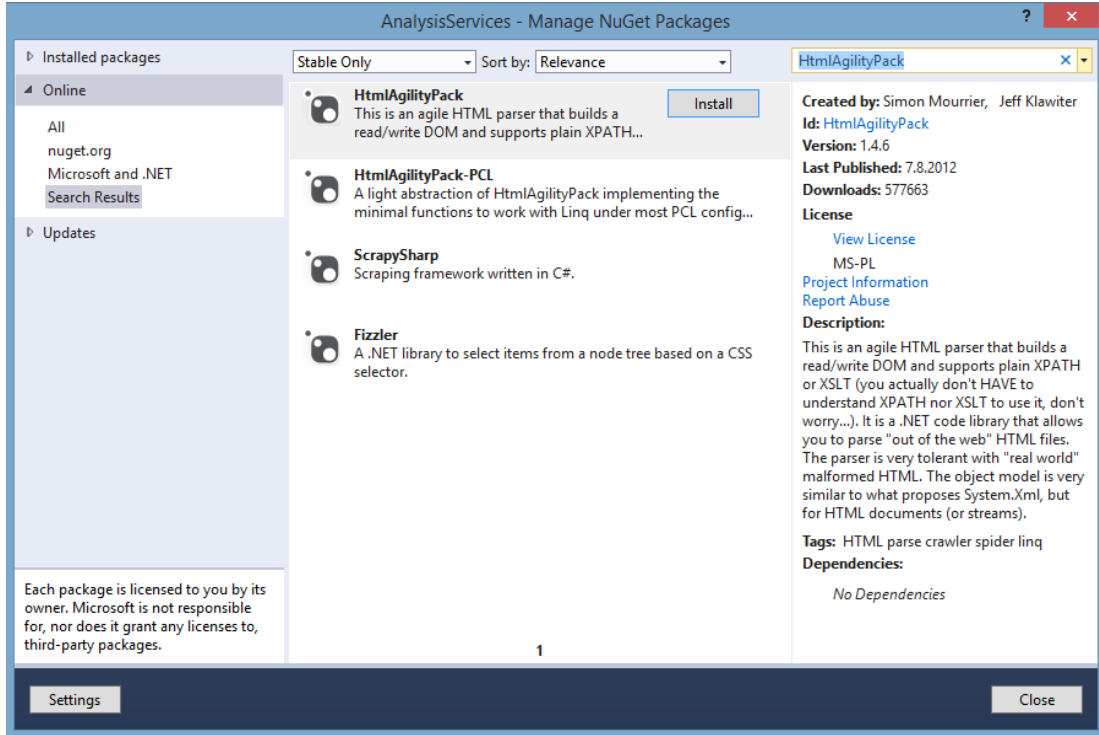
4.4.1 Veri Toplama

Bu aşamada daha önce de bahsedildiği gibi veri ambarı uygulaması kullanılacaktır. Uygulama içerisinde sikayetvar.com gibi sosyal paylaşım sitelerinden kullanıcı yorumları toplanıp veritabanına eklenir. Bu işlemi yaparken sikayetvar.com üzerindeki yorumları okuyabilmek için HtmlAgilityPack.dll kullanılır. Bu sınıfı uygulamaya ekleyebilmek için Şekil 4.5'te görüldüğü gibi Manage Nuget Packages uygulaması kullanılır. Bunun için öncelikle projeyi sağ tıklayıp Manage Nuget Packages tıklanır.



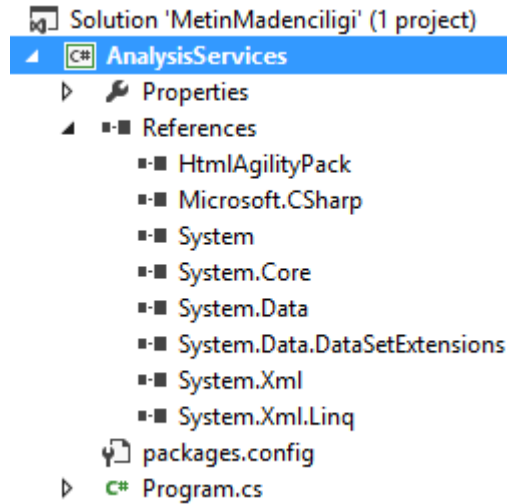
Şekil 4.5 Nuget uygulamasına ulaşma

Daha sonra açılan ekrandan search alanına HtmlAgilityPack yazılır ve enter tuşuna basılır, gelen sonuçtan Şekil 4.6'da görüldüğü gibi Install butonuna basılır.



Şekil 4.6 Class ekleme adımı

Uygulamanın referans bölümünde seçilen sınıfın eklendiği Şekil 4.7’de görülmektedir.



Şekil 4.7 Uygulamanın Referans Listesi

Bu işlemleri yaptıktan sonra sikayetvar.com sitesi üzerinden verileri okuyarak, entity framework teknolojisini kullanarak, MSSQL Server üzerinde tutulan veritabanına kaydetme işlemini yapacak olan metodları hazırlanır. Öncelikle sikayetvar.com üzerinden veriyi çekecek metodlar görülür.

```

1 reference
public static void BaslikYaz()
{
    for (int x = 0; x < 100; x++)
    {
        Uri url = new Uri("http://www.sikayetvar.com/arama/burger%20king/s/" + x);
        WebClient client = new WebClient();
        string html = client.DownloadString(url);
        HtmlAgilityPack.HtmlDocument htmldoc = new HtmlAgilityPack.HtmlDocument();
        htmldoc.LoadHtml(html);
        var basliklar = htmldoc.DocumentNode.SelectNodes("/html/body/div[1]/div/div/ul/li/div/span");
        List<string> liste = new List<string>();
        foreach (var baslik in basliklar)
        {
            liste.Add(baslik.InnerText);
        }
        for (int i = 0; i < liste.Count; i++)
        {
            DetayYaz(liste[i].Split('|')[2].TrimEnd().TrimStart(), Convert.ToDateTime(liste[i].Split('|')[1]));
        }
    }
}

```

Şekil 4.8 Başlık Yaz Metodu

Şekil 4.8'de ki BaslikYaz metodu site üzerinde işlenmesi istenen firma için arama işlemi yapar ve çıkan sonuçları tek tek dolaşarak detay linklerini detay göster metoduna iletir.

```

1 reference
public static void DetayYaz(string ID, DateTime DT)
{
    Uri url = new Uri("http://www.sikayetvar.com/sikayet/detay/" + ID);
    WebClient client = new WebClient();
    string html = client.DownloadString(url);
    HtmlAgilityPack.HtmlDocument htmldoc = new HtmlAgilityPack.HtmlDocument();
    htmldoc.LoadHtml(html);
    var basliklar = htmldoc.DocumentNode.SelectNodes("/html/body/div[1]/div/div/div[1]/p");
    List<string> liste = new List<string>();
    try
    {
        foreach (var baslik in basliklar)
        {
            liste.Add(ReplaceText(baslik.InnerText.TrimEnd().TrimStart()));
        }

        foreach (var item in liste)
        {
            using (SocialMediaDBEntities EF = new SocialMediaDBEntities())
            {
                EF.TextAdd(1, item, DT);
            }
        }
    }
    catch (Exception)
    { }
}

```

Şekil 4.9 Detay Yaz Metodu

Şekil 4.9’da gördüğümüz DetayYaz metodu BaslikYaz metodundan aldığı detay linkine gider ve o sayfada ki yorum satırını alıp Entity Framework üzerinde bulunan TextAdd fonksiyonunu kullanarak veritabanına kayıt işlemini gerçekleştirir.

```

1 reference
public static string ReplaceText(string _text)
{
    _text = _text.Replace("Ä°", "i")
    .Replace("Ä±", "ı")
    .Replace("Ä&frac14;", "ü")
    .Replace("Äÿ", "ş")
    .Replace("Ä", "ş")
    .Replace("Ä§", "ç")
    .Replace("Ä¶", "ö")
    .Replace("Äÿ", "ğ")
    .Replace("Ä‡", "ç")
    .Replace("Ä-", "ö")
    .Replace("Äœ", "ü")
    .Replace("Ä%", "ü")
    .Replace("&#039;", "'")
    .Replace("?", "");
    return _text;
}

```

Şekil 4.10 ReplaceText Metodu

ReplaceText metodu ise DetayGoster metodu üzerinde okunmuş yorumlar üzerinde karakter sınıfı farklılığından ötürü oluşan karakter hatalarını düzeltme işlemi yapmaktadır.

4.4.2 Veri Temizleme ve Kategori Oluşturma

Öncelikle toplanılan veriler içerisinde algoritmayı geciktirecek, yoracak ve sonuçları etkileyecek veriler temizleme işlemine tabii tutuldu. Bir gıda uzmanının görüşü doğrultusunda temizlenmiş yorumlara bakılarak analiz kategorileri belirlenip veri tabanına yazıldı. Bu tezin uygulama bölümünde kategoriler Tablo 4.1’de görülmektedir.

Tablo 4.1 Kategoriler

Kategori Adı	Açıklama
Hijyen	Yiyecek, içecek ve mağaza temizliğini kapsar.
Hız	Verilen siparişin ne kadar sürede hazır edildiği bilgisini inceler.
Lezzet	Üretilen ürünlerin lezzetinin beklentileri ne kadar karşıladığını inceler.
Memnuniyet	Genel anlamda çalışanların davranışlarını kapsar.

4.4.3 Gövdeleme

Uygulama çalışmasında gövdeleme yöntemi olarak Joker (Wildcard) yöntemi kullanılmıştır. Bu yöntemde, kelimelerin kendileri değil gövdeleri kullanılır. Çünkü Türkçe ve benzeri dillerde bir kelime sonuna birden fazla ek alarak birçok farklı anlamlar kazanabilmektedir. Örneğin "kıl" kelimesi ile "kıllı", "kıllıydı" kelimeleri joker yöntemi kullanılmazsa her biri farklı bir kelime olarak görülecekti ve bunlar sözlüğe ayrı ayrı kaydedilecekti. Bu duruma bağlı olarak sözlük boyutu artacak ve sınıflandırma aşamasında başarı düşecekti. Yine sözlük boyutuna bağlı olarak algoritmanın işlem süresi artacak ve uygulama geç cevap verecekti.

Joker kelime, aynı kök dizini ile başlayan ve çeşitli ekler almış ancak anlam olarak birbirine yakın anlamda olan kelimeleri tek bir grup altına toplayan kelimelerdir. Eklerden ayrıştırılan kelimeler ortak köke indirgenir (Pilavcılar, 2007).

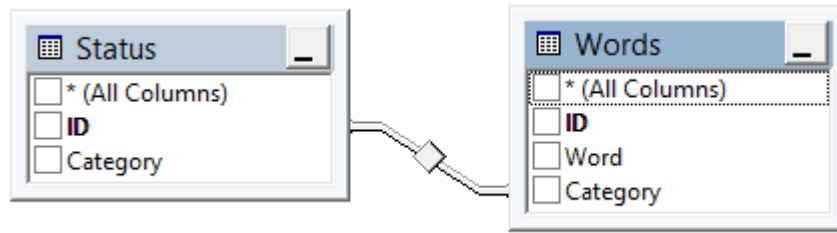
4.4.4 Sözlük Oluşturma

Bu aşamada oluşturulan kategorilere uygun olarak bir sözlük oluşturuldu. Bu noktada dikkat edilmesi gereken oluşturulan sözlükte gereksiz kelimelere yer vermemektir. Çünkü bu gibi durumlar sonuçları olumsuz etkileyebilmektedir. Çoğu zaman en sık karşılaşılan kelimelerden oluşturulmuş sözlükler çoğu zaman başarılı sonuçlar vermektedir. Gıda sektörünün şikâyetlerini temsil edebilecek dört kategori altında oluşturulan sözcükler Tablo 2'de görülmektedir. Bu sözcüklerin belirlenmesinde de bir gıda uzmanının görüşü alınmıştır.

Tablo 4.2 Kategoriler ve Sözcükler

Hijyen	Hız	Lezzet	Memnuniyet
Pis	Yavaş	Kötü	Saygısız
Tuvalet	Ağır	İğrenç	Terbiyesiz
Çamur	Bekle	Soğuk	Eksik
Kıl	Geç	Sert	İlgisiz
Kan	Bekliyor	Bayat	Kaba
Saç	Saat	Buz	Amatör
Eldiven	Dakika	Lezzetsiz	Yanlış
Tırnak	Gelmedi	Donuk	Suratsız

Sözlük yapısı veri tabanı üzerinde şu şekilde saklanmaktadır.

**Şekil 4.11 Uygulama Sözlük Yapısı**

Veri tabanı üzerinde bulunan Status ve Words tabloları Status. ID, Words. Category alanları üzerinden birbirleriyle ilişkilidir. Bu ilişkiler üzerinden ayrıştırılmış kelimeler, veriler üzerinde kategori bazında uygulanmaktadır.

4.5 ALGORİTMA UYGULAMA

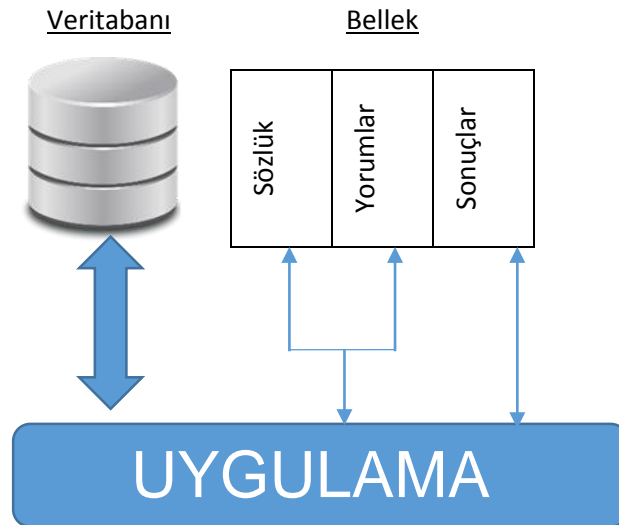
Veri toplama işlemi yaptıktan sonra sözlük oluşturulmuş ve bu veriler üzerinden algoritma uygulayarak raporlama işlemleri yapılmıştır.

Birinci aşamada sözlükten gruplarına göre kelimeler bellek üzerine yazılır.

```
List<WordListResult> Hijyen = DB.WordList(1).ToList();
List<WordListResult> Hiz = DB.WordList(2).ToList();
List<WordListResult> Lezzet = DB.WordList(3).ToList();
List<WordListResult> Memnuniyet = DB.WordList(4).ToList();
```

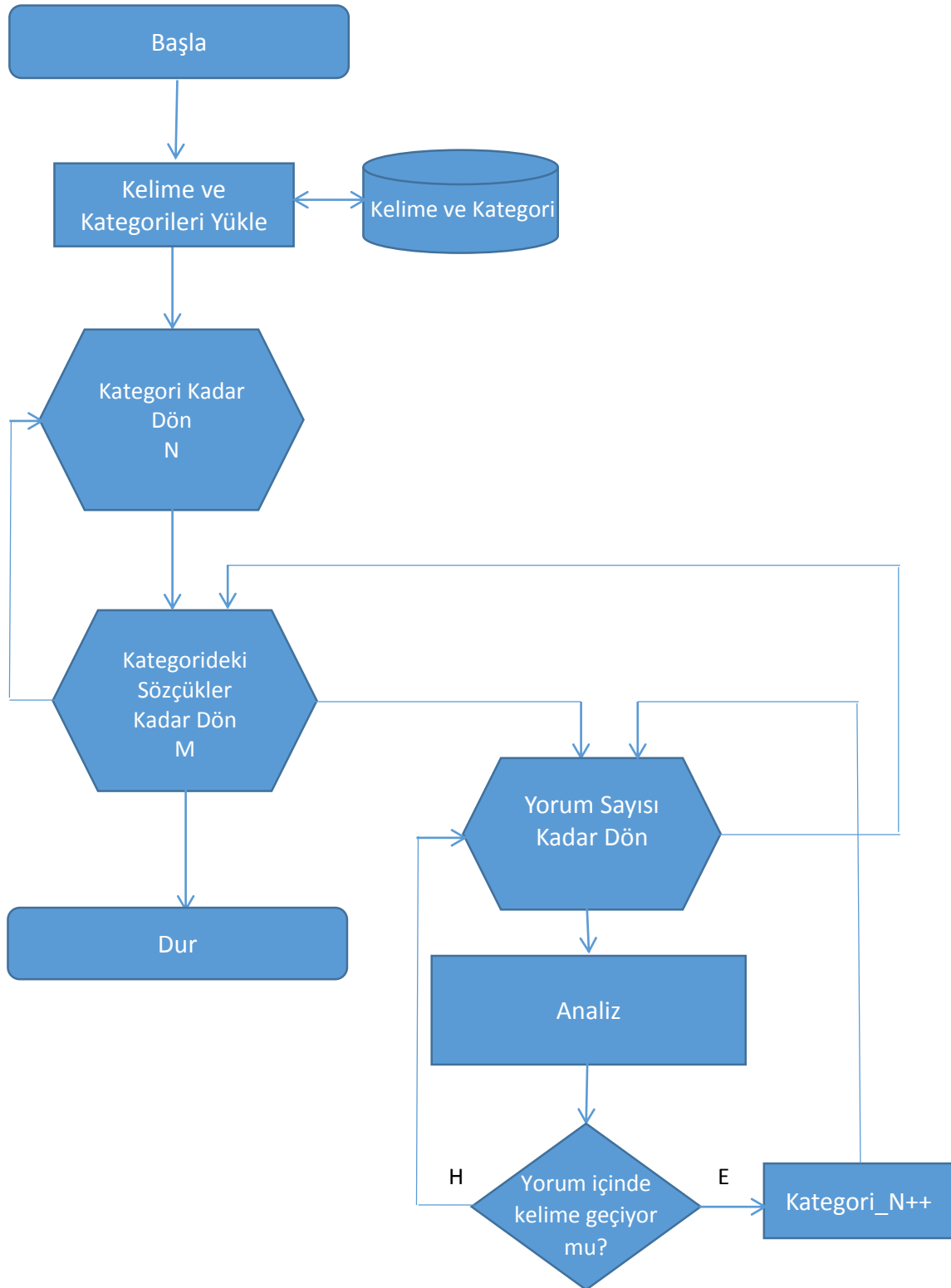
Şekil 4.12 Sözlük Bilgisi Çekme

Daha sonra seçilen firmanın veritabanı üzerinden yorumları çekilir ve bu yorumları belleğe, yazılan sözlük kelimelerinden grup bazında analize sokulur. Yapılan her analizde sonuçlar bir liste içeresine atılır ve bu liste bir raporlama aracına bağlanır. Şekil 4.13'te işlemler şema üzerinde görülmektedir.



Şekil 4.13 Uygulama Şeması

Şekil 4.14'te uygulama şemasında bulunan uygulama alanının algoritma işleyişi gösterilmektedir.



Şekil 4.14 Metin Madenciliği Algoritmasının Akış Diyagramı

Örnek olarak “A firmasından eve sipariş verdik, yanlış sipariş göndermişler. Arayıp tekrar istedim, yemek zaten soğuk gelmişti, iyice soğudu. Ayrıca son zamanlarda A firması her şeyini bozdu! Kendinizi bir an önce düzeltin!” yorumunu ele alalım. Sözlüğümüzde bulunan “yanlış” ve “soğuk” kelimeleri cümlelerin içinde geçmektedir. Yanlış” kelimesinin bulunduğu kategori memnuniyetsizlik kategorisidir. “Soğuk” kelimesi ise lezzetsizlik kategorisi altına girmektedir. Yani bu yorumda hem memnuniyetsizliğe hem de lezzetsizliği atıfta bulunuluyor. Tablo 4.3’de görüldüğü gibi ilgili kategorilerin sayaçları artırılır. Daha sonra her bir kategorinin sayacı yorumlarda geçen sözlükle uyumlu toplam kelime sayısına bölünerek her bir kategorinin yüzdesi bulunur.

Tablo 4.3 Yorum 1 sonrası analiz sonucu

Yorum	Hijyen	Hız	Lezzet	Memnuniyet
Yorum 1			X	X
Toplam	0	0	1	1
Yüzde	%0	%0	%50	%50

“Fenerbahçe stadyumunun oradaki A firmasından çok ama çok şikayetçiyiz. Her geldiğimizde 25 dk bekliyoruz. Her geldiğimizde. Bu ne biçim hizmet kalitesi” yorumunda ise bekliyoruz kelimesi bizim içi anahtar kelime olarak alınıyor. Sözlüğümüzde “bekle” ve “bekliyor” şeklinde iki kelime hız kategorisi altında tutulmaktadır. Bekle kelimesi bekliyoruz kelimesi ile eşleşmeyecektir. Çünkü kelime kökenleri farklılık oluşturmaktadır. Fakat “bekliyor” kelimesi üzerinden hız kategorisi altında eşleşme sağlanacaktır. Yorum 2 sonrası analiz sonucu Tablo 4.4’de görülmektedir.

Tablo 4.4 Yorum 2 sonrası analiz sonucu

Yorum	Hijyen	Hız	Lezzet	Memnuniyet
Yorum 1			X	X
Yorum 2		X		
Toplam	0	1	1	1
Yüzde	%0	%33,3	%33,3	%33,3

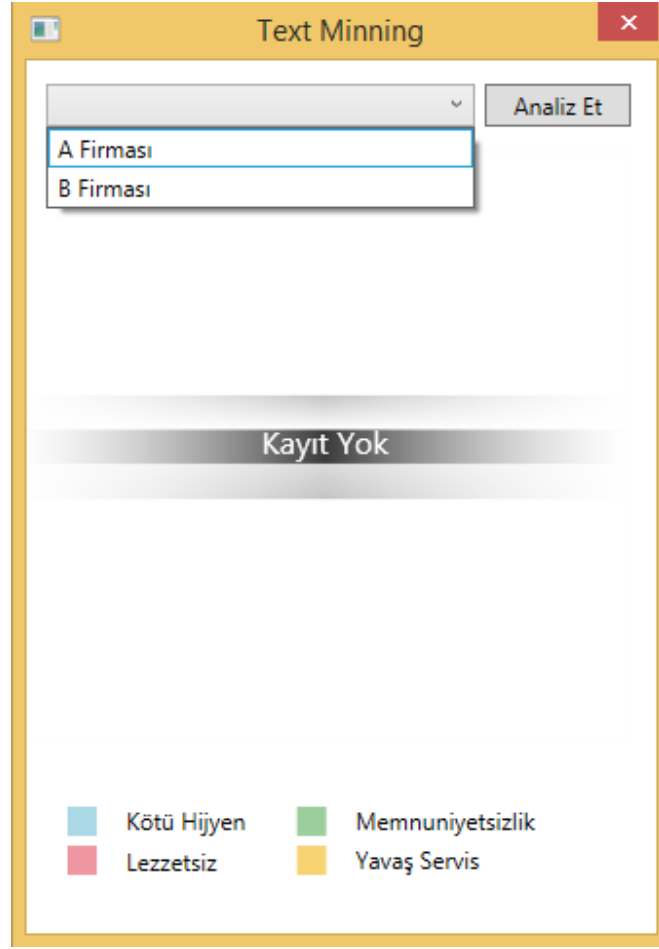
“A firması Sefaköy şubesi 10 dakika mesafede olmama rağmen siparişimi getirmede. Yolda olan siparişim amatör yol bilmeyen eleman yüzünden iptal oldu getirmediler. Buradan asla sipariş vermeyeceğim, ayrıca daha önce pis oldukları gerekçesiyle arkadaşlarımdan şikayet almıştım inanmadım fakat bu kadarı da fazla. Yemeksitesinden sipariş vermeme rağmen siparişim gelmedi. Yemeksitesinin görevi nedir” yorumunda üç kelime dikkatimizi çekiyor. İlk olarak “dakika” kelimesi ile hız kategorisine atıfta bulunuluyor. İkinci kelitemiz “pis” kelimesi yemeklerin hijyen konusunda başarısız olduğu belirtiliyor ve üçüncü kelitemiz ise “amatör” kelimesi ile memnuniyetsizlik belirtiliyor. Analiz sonuçları Tablo 4.5’de verilmiştir.

Tablo 4.5 Yorum 3 sonrası analiz sonucu

Yorum	Hijyen	Hız	Lezzet	Memnuniyet
Yorum 1			X	X
Yorum 2		X		
Yorum 3	X	X		X
Toplam	1	2	1	2
Yüzde	%16,6	%33,3	%16,6	%33,3

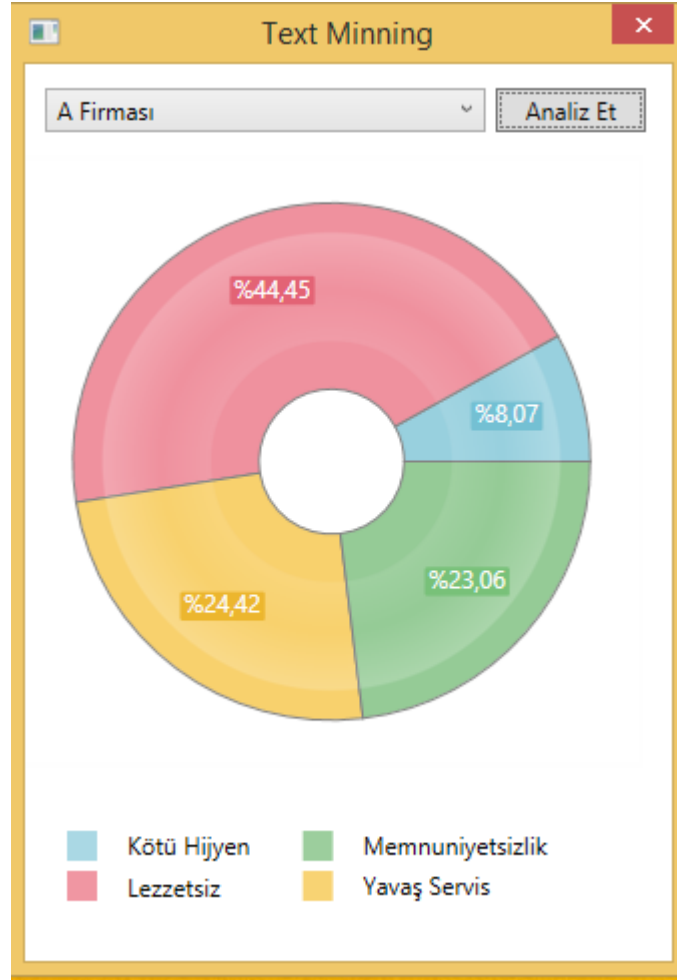
Örneklere görüldüğü gibi sözlükte ki her kelime yorumlar içerisinde aranır. Bulunan kelimelerin bulunduğu kategoriye bakılır ve yoruma bu kategori sınıfı için işaret konulur.

Yorumlar işaretlenme sayısına göre bir raporlama aracına verilir ve ekranda gösterilir. Şimdi bu algoritmayı işleten programımızın arayüzünü inceleyelim.



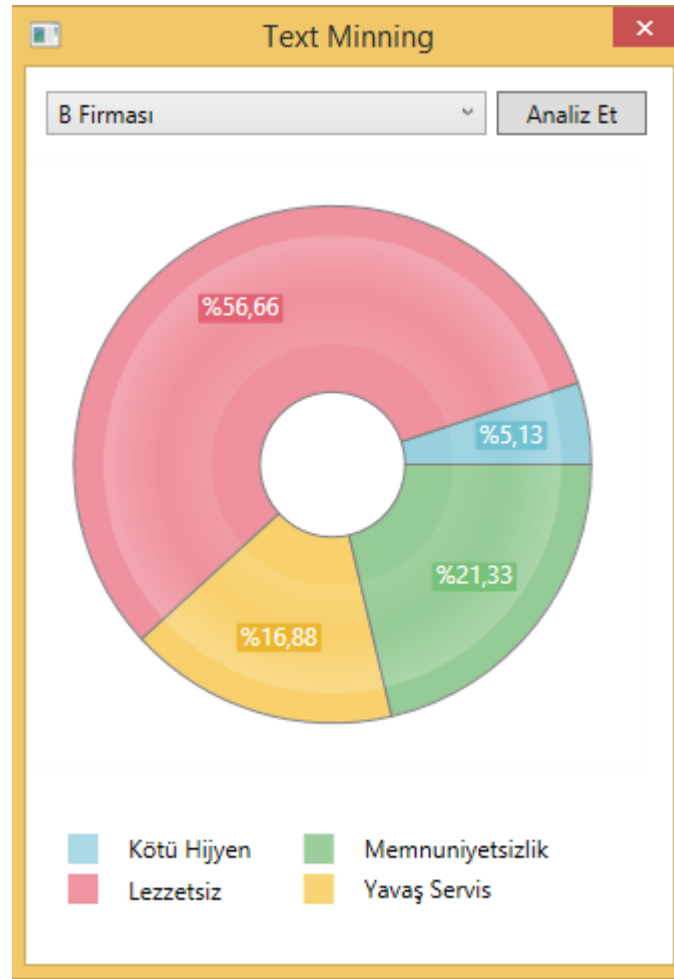
Şekil 4.15 Analiz Programı Raporlama Arayüzü - 1

Şekil 4.15 görüldüğü gibi bir combobox içerisinde analiz yapılacak şirket isimleri bulunmaktadır. Buradan analiz edilmek istenilen şirketin ismini seçtikten sonra “Analiz et” butonuna basılır ve raporlama sürecini bekledikten sonra ekranda Chart üzerinde raporlama görünür.



Şekil 4.16 Analiz Programı Raporlama Arayüzü - 2

İlk şirket için analiz yapıldığında görüldüğü gibi belirlenen dört kategori üzerinden bir rapor hazırlanmış ve yüzdelik oranlar üzerinden chart kullanılarak ekrana yansıtılmıştır. Bundan sonraki bölümde yapılan analiz diğer şirketin sonuçları sunulmaktadır.



Şekil 4.17 Analiz Programı Raporlama Arayüzü - 3

İkinci şirket için yapılan analiz ile ilk şirket için yapılan analiz oranlarda önemli bir farklılık olduğu görülmektedir.

Bu iki rapor üzerinden iki şirketin birbirlerinden üstün veya başarısız yanları ortaya çıkarılmaktadır. Buna örnek vermek gerekirse ikinci şirket, lezzet açısından %56,66 oranında şikayet almış, fakat ilk şirket %44,45 oranında şikayet almıştır. Bu oranlara dayanarak lezzet konusunda birinci şirketin daha başarılı olduğu görülmektedir. Diğer bir kategori olan yavaş servis oranlarına bakıldığında ilk şirketin aldığı şikayetlerin %24,42'sini "yavaş servisin" oluşturduğu görülmektedir, fakat ikinci şirketimiz %16,88 oranında şikayet almış ve hız konusunda birinci şirketi geride bırakmıştır.

5. SONUÇ

Günümüzde müşteri memnuniyetinin ön planda olduğu bir pazarda onların beklentilerini bilmek ve memnuniyetini sağlamak firmaların en önemli ilkelerindedir. Müşteri beklentilerini tahmin edebilmek için firmalar büyük bütçeler harcamaktadır. Fakat bu harcamalar yeterli insana ulaşamadığı zamanlar başarısız sonuçlar doğurabilmektedir ve ciddi bir zaman gerektirmektedir.

Son yıllarda başarılı olan firmalar incelendiğinde sosyal medyayı aktif olarak kullandıklarını görülmektedir. İnternet üzerinde yapılan kısa bir araştırmada insanlar sosyal medya üzerinde yaptıkları paylaşımlar ve yorumlar ile ilgili mutlaka hizmet teklifleri aldıklarını söylemektedirler. Bu da firmaların artık sosyal medya üzerinde müşteri kazanabilmek için büyük bir havuz olduğunun farkında olduklarını göstermektedir.

Firmalar artık sosyal medya kullanarak insanların beklentilerini daha az harcama ile daha çok insana ulaşarak öğrenebilmektedirler. Metin Madenciliği kullanılarak sosyal medya üzerinden yapılan yorumlar ile firmalar sundukları hizmetlerin insanlar üzerinde ki başarısını veya başarısızlıklarını çok rahat ölçebilmektedir. Başarısız olduklarını gördükleri konulara daha ağırlık vererek hizmetlerinin kalitesini arttırabilirler. Yine sosyal medya üzerinden insanların beklentilerini gözlemleyerek hizmet anlayışlarını analiz sonuçlarına göre şekillendirebilirler.

Tezin uygulama bölümünde görüldüğü gibi bir firma için uygulamanın sunduğu analizler oldukça yönlendirici veriler oluşturmuştur. Bu verilere göre müşterilerinin en yoğun şikâyette bulunduğu konuları keşfedebilir ve bu konuları ele alarak çözüm yoluna gidilebilir. Bu yöntem hem ucuz hem de daha çok insana hitap eden bir araştırma sonucu oluşturacaktır. Buna ek olarak bu çalışma bir program aracılığıyla gerçekleştirileceği için zaman konusunda da kazançlı olunacaktır.

Sosyal medya dışında giriş bölümünde bahsedildiği gibi metin madenciliği yöntemi “ürün puanlama” yapmak için kullanılabilir. Yapılan yorumlar amaca yönelik hazırlanan bir sözlük içeriğinden geçirilerek ürün bazında insanlara belirleyici bilgiler gösterilebilir.

Bir başka kullanım alanı olarak iş başvurularında CV’ler üzerinden bilgi toplama işlemleri gösterilebilir. Aranılan kriterlere en uygun kişi metin madenciliği kullanılarak birçok CV arasından filtrelenebilir. Metin madenciliği reklamcılık alanında aktif olarak kullanılabilir. Örneğin elinizdeki birçok reklam içeriğinden kişiye en uygun reklam web sitesi üzerinden gösterilebilir. Kişilerin yaptığı aramalar veya incelediği ürünler bu işlem için belirleyici bir veri olacaktır. Makale içerikli web siteleri üzerinde yazının içeriği yine bir sözlük filtresinden geçirilerek otomatik kategorileme işlemleri yapılabilmektedir. Bu yöntem kullanılarak doküman kategorileme işlemleri de yapılabilir.

Metin madenciliği birçok alanda aktif olarak kullanıldığı gibi sosyal medya da oldukça verimli sonuçlar vermektedir ve firmaların bu teknolojiye önümüzdeki yıllarda daha da aktif olarak yararlanmaları beklenmektedir. Firmalara bu analizlerden aldıkları raporlar ile son kullanıcıya yönelik, hatta kişi bazında daha kaliteli, verimli ve kişi için anlam ifade eden hizmetler vereceği öngörülebilir.

Metin madenciliği yöntemlerinin maliyetinin az, verimin yüksek olduğu düşünülürse kullanımının gelecek yıllarda daha çok olacağı, kullanım alanlarının da bu gelişmeye bağlı olarak artacağını ve bazı meslek dalları için vazgeçilmez bir raporlama ve analiz yöntemi olacağı söylenebilir.

KAYNAKÇA

- Akpınar H. (2000). *Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği*, İ.Ü. İşletme Fakültesi Dergisi, İstanbul, Sayı:1,Nisan 2000)
- Alpaydın E. (2000). *Zeki Veri Madenciliği, Ham Veriden Altın Bilgiye Ulaşma Yöntemleri*". *Bilişim 2000 Eğitim Semineri, Boğaziçi Üniversitesi Bilgisayar Mühendisliği Bölümü.*
- Aşlıyan R. Ve Günel K. (2010). *Metin İçerikli Türkçe Dokümanların Sınıflandırılması*
- Bhatt C. (2004). *Mining the Medical Literature.*
[http://ai.stanford.edu/~serafim/CS374_2004/Lecture%20Notes/lecture6.pdf.](http://ai.stanford.edu/~serafim/CS374_2004/Lecture%20Notes/lecture6.pdf)
- Bilgi Türleri.(1986). *Büyük Larousse Sözlük Ve Ansiklopedisi*, (Cilt 23, s. 12164- 12165). İstanbul, İnterpress Basın ve Yayıncılık.
- Buss K. (2011). *Literature Review on Preprocessing for text minning*
[http://dmu-ca.ioct.dmu.ac.uk/publication/papers/literature_review_kono.pdf.](http://dmu-ca.ioct.dmu.ac.uk/publication/papers/literature_review_kono.pdf)
- Eker H. (2005). *Veri Madenciliği ve Bilgi Keşfi.*
- Ergün K. (2012). *Metin Madenciliği Yöntemleri İle Ürün Yorumlarının Otomatik Değerlendirilmesi*
- Feldman R. ve Sanger J. (2007). *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, USA.*
- Gürsoy Ş. U. T. (2010) , *Uygulamalı Veri Madenciliği*. Ankara: Pagem Akademi
- Karaca, M. F. (2012). *Metin Madenciliği Yöntemi İle Haber Sitelerindeki Köşe Yazılarının Sınıflandırılması*
- Karadağ A. Ve Takçı H. (2010). *Metin Madenciliği ile Benzer Haber Tespiti*

- Kayabaşı, A. (2010). *Elektronik (online) Alışverişte Lojistik Faaliyetlere Yönelik Müşteri Şikayetlerinin Analizi ve Bir Alan Araştırması. İşletme Araştırmaları Dergisi,41-42.*
- Konchady M. (2006). *Text Mining Application Programming. 1st ed. Charles River Media.*
- Miller, T. W. (2005). *Data and Text Mining A Business Applications Approach, Pearson Education Inc, USA.*
- Munyaradzi C. ve Al-Ayyoub M. ve Hossain M.S. ve Gupta R. (2011). *CSE 634-Data Mining: Text Mining*
- Oğuzlar, A. (2011). *Temel Metin Madenciliği.* Bursa: Dora Yayıncılık Ltd. Şti.
- Özkan, Y. (2008) . *Veri Madenciliği Yöntemleri.* İstanbul: Papatya Yayıncılık.
- Pilavcılar, İ. F. (2007). *Metin Madenciliği ile Metin Sınıflandırma,* Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Uzunköprü, S. (2014). *Projeler ile C# 5.0 ve SQL SERVER 2012.* İstanbul: Kodlab Yayıncılık.
- Weiss, S. M. ve Indurkha, N. ve Zhang, T. ve Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information.* Springer.
- Wu H. ve Shenghua Z. ve Ling L. (2012). *Social media competitive analysis and text mining: A case study in the pizza industry*
- Yağmur O. (2009) *WPF'i Anlamak*

WEB KAYNAKLARI

- [1] http://ai.stanford.edu/~serafim/CS374_2004/Lecture%20Notes/lecture6.pdf
- [2] <http://bilgiagaci.com>
- [3] <http://cogsci.princeton.edu/cgi-bin/webwn>
- [4] <http://csharpdevelopers.net>
- [5] <http://emraharslanbm.wordpress.com/>
- [6] <http://hlst.sabanciuniv.edu/TL>
- [7] <http://tuvakitaplar.tumblr.com>
- [8] <http://wvf.nedir.com/>
- [9] (<https://www.ling.upenn.edu>)

EKLER

EK-1: Siteden başlıkları ve detay linkini okuyan Class

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Net;
using System.Text;

namespace DataWareHouse
{
    public class HTMLRead
    {
        public static void BaslikYaz()
        {
            for (int x = 0; x < 100; x++)
            {
                Uri url = new Uri("http://www.sikayetvar.com/arama/afirmasi/s/" +
x);

                WebClient client = new WebClient();
                string html = client.DownloadString(url);
                HtmlAgilityPack.HtmlDocument htmldoc = new
HtmlAgilityPack.HtmlDocument();
                htmldoc.LoadHtml(html);
                var basliklar =
htmldoc.DocumentNode.SelectNodes("/html/body/div[1]/div/div/ul/li/div/span")
;

                List<string> liste = new List<string>();
                foreach (var baslik in basliklar)
```

```

    {
        liste.Add(baslik.InnerText);
    }
    for (int i = 0; i < liste.Count; i++)
    {
        DetayYaz(liste[i].Split('|')[2].TrimEnd().TrimStart(),
Convert.ToDateTime(liste[i].Split('|')[1]));
    }
}
}

public static void DetayYaz(string ID, DateTime DT)
{
    Uri url = new Uri("http://www.sikayetvar.com/sikayet/detay/" + ID);
    WebClient client = new WebClient();
    string html = client.DownloadString(url);
    HtmlAgilityPack.HtmlDocument htmldoc = new
HtmlAgilityPack.HtmlDocument();
    htmldoc.LoadHtml(html);
    var basliklar =
htmldoc.DocumentNode.SelectNodes("/html/body/div[1]/div/div/div[1]/p");
    List<string> liste = new List<string>();
    try
    {
        foreach (var baslik in basliklar)
        {
            liste.Add(ReplaceText(baslik.InnerText.TrimEnd().TrimStart()));
        }

        foreach (var item in liste)
        {

```

```

        using (SocialMediaDBEntities EF = new
SocialMediaDBEntities())
        {
            EF.TextAdd(1, item, DT);
        }
    }
}
catch (Exception)
{
}
}
}

```

```

public static string ReplaceText(string _text)
{
    _text = _text.Replace("Ä°", "ı")
        .Replace("Ä±", "ı")
        .Replace("Ä&frac14;", "ü")
        .Replace("Äÿ", "ş")
        .Replace("Ä", "Ş")
        .Replace("Ä§", "ç")
        .Replace("Ä¶", "ö")
        .Replace("Äÿ", "ğ")
        .Replace("Ä‡", "Ç")
        .Replace("Ä–", "Ö")
        .Replace("Äœ", "Ü")
        .Replace("Ä¼", "ü")
        .Replace("&#039;", "")
        .Replace("?", "");
    return _text;
}
}

```



```
}  
}
```

EK-2: Raporlama işlemini yapıldığı class

```
using System;  
using System.Collections.Generic;  
using System.Linq;  
using System.Text;  
using System.Windows;  
using System.Windows.Controls;  
using System.Windows.Data;  
using System.Windows.Documents;  
using System.Windows.Input;  
using System.Windows.Media;  
using System.Windows.Media.Imaging;  
using System.Windows.Navigation;  
using System.Windows.Shapes;  
using Telerik.Charting;  
using Telerik.Windows.Controls.Charting;
```

```
namespace WpfApplication1  
{  
    public partial class MainWindow : Window  
    {  
        DbContext DB = new DbContext();  
        public MainWindow()  
        {  
            InitializeComponent();  
        }  
    }  
}
```

```

btnAnaliz.Click += btnAnaliz_Click;
cmdSirket.DisplayMemberPath = "ScreenName";
cmdSirket.SelectedValuePath = "ID";
cmdSirket.ItemsSource = DB.ScreenNameList();
}

```

```

void btnAnaliz_Click(object sender, RoutedEventArgs e)

```

```

{

```

```

    try

```

```

    {

```

```

        List<WordListResult> Hijyen = DB.WordList(1).ToList();

```

```

        List<WordListResult> Hiz = DB.WordList(2).ToList();

```

```

        List<WordListResult> Lezzet = DB.WordList(3).ToList();

```

```

        List<WordListResult> Memnuniyet = DB.WordList(4).ToList();

```

```

        ScreenNameListResult result =

```

```

        (ScreenNameListResult)cmdSirket.SelectedItem;

```

```

        List<ChartDataItem> Liste = new List<ChartDataItem>();

```

```

        int a = 0;

```

```

        foreach (var obj in Hijyen)

```

```

        {

```

```

            foreach (var item in DB.TextList(result.ID))

```

```

            {

```

```

                if (item.Text.IndexOf(obj.Word) != -1)

```

```

                {

```

```

                    a++;

```

```

                }

```

```

    }
}

Liste.Add(new ChartDataItem() { Aciklama = "Kötü Hijyen", Deger
= a });

int b = 0;
foreach (var obj in Hiz)
{

    foreach (var item in DB.TextList(result.ID))
    {

        if (item.Text.IndexOf(obj.Word) != -1)
        {
            b++;
        }
    }
}
Liste.Add(new ChartDataItem() { Aciklama = "Lezzetsiz", Deger = b
});

int c = 0;
foreach (var obj in Lezzet)
{

    foreach (var item in DB.TextList(result.ID))
    {

        if (item.Text.IndexOf(obj.Word) != -1)
        {
            c++;
        }
    }
}

```

```

    }

    Liste.Add(new ChartDataItem() { Aciklama = "Memnuniyetsizlik",
Deger = c });

    int d = 0;
    foreach (var obj in Memnuniyet)
    {
        foreach (var item in DB.TextList(result.ID))
        {
            if (item.Text.IndexOf(obj.Word) != -1)
            {
                d++;
            }
        }
        Liste.Add(new ChartDataItem() { Aciklama = "Yavaş Servis", Deger
= d });
        chart2.ItemsSource = Liste;
    }
    catch (Exception)
    {
        MessageBox.Show("Değer Seçiniz", "Hata",
MessageBoxButton.OK, MessageBoxImage.Error);
    }
}

public class ChartDataItem
{
    public string Aciklama;

```

```
    public int Deger;  
  }  
}
```

ÖZET

Teknolojinin gelişmesi ve toplumlarda yaygınlaşması ile artık internet kullanımı da artmış bulunmaktadır. Bu artış internet ortamında ciddi boyutlarda kaynaklar oluşmasını sağlamıştır. Bu kaynakların belli bir bölümü işlenmiş belli bir bölümü ise ham kaynaklardır. Bu ham kaynaklar işlendiği zaman belki de internet ortamından başka hiçbir ortamda ulaşılamayacak devasa raporlar ortaya çıkarmaktadır.

Bu büyük boyutlu kaynakları işlemek ciddi çalışmalar ve yöntemler gerektirmektedir. Bu yöntemlerden biri de metin madenciliğidir. Metin madenciliği günümüzde hızla yaygınlaşan ve kullanışlı bir veri işleme yöntemidir. Bu çalışmada, ham kaynakların sosyal medya üzerinden nasıl toplanacağını, firmalar açısından metin madenciliği kullanarak nasıl işleneceğini ve anlamlı hale gelen bu verilerin nasıl raporlanacağı müşteri şikayet verileri üzerinden anlatılmaktadır. Bu çalışma, müşteri kayıplarını önleme ve şirket karını artırmak açısından yöneticiler için önem arz etmektedir.

Anahtar Sözcükler

1. Metin Madenciliği
2. Veri Madenciliği
3. Sosyal Medya
4. Müşteri Şikayet Analizi

ABSTRACT

Internet usage has increased by development of technology and diffusion in societies. This increase has provided to create sources in serious dimensions on the internet environment. Some of these sources are processed and some of them are raw sources. When these raw sources are processed, it has probably presented huge reports which cannot be reached anywhere from the internet environment.

Processing of these big sized resources has required serious studies and methods. One of these methods is text mining. The text mining is a data processing method which is useful and become widespread rapidly. In this study, how to collect the raw materials on the social media, how to process them with respect to firms by using the text mining and how to report these data becoming significant are explained on customer complaint data. This study is valuable for managers to avoid the customer loss and to increase the profit of their firms.

Key Words

1. Text Mining
2. Data Mining
3. Social Media
4. Customer Complaint Analysis

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : Gökhan Aravi
Telefon : 0555 316 76 76
E-Mail : gokhanaravi@gmail.com
Web Sitesi : www.gokhanaravi.com

Genel Bilgiler

Eğitim Durumu :
Üniversite-Yüksek Lisans
İstanbul Aydın Üniversitesi – Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği (Tezli)

Medeni Durumu : Bekar
Doğum Yeri ve Tarihi : 14.12.1986 – İstanbul
Ehliyet : B
Kan Grubu : 0 RH+
Yabancı Dil : İngilizce

Projeler

SUPERONLINE FTTX DASHBOARD PROJESİ

Supeornline ait Türkiye üzerinde ki switchlerin anlık olarak ölçümlerini Ajax teknolojisi ile hazırlanmış WebArayüzde gösterilmesi ve Down'dan up'a dönen switchlerin service desk uygulaması üzerinden takibi yapılması up olan switch ticket'ı kapatılması.(3 version geliştirildi)

SUPERONLINE ZİNCİR PROJESİ

Maximo ve Service Desk uygulamaları arasında manuel olarak gerçekleştirilen iş sürecini sistem üzerinde otomatik bir yapıya dönüştürerek iki uygulama arasında haberleşme ve iş akış sürecini manuel bir senkronizasyon yapmadan ilerletilmesi.

TURKCELL MAXIMO VE SD TICKET ENTEGRASYON GELİŞTİRİLMESİ

Maximo ve Service Desk uygulamaları arasında Webservice aracılığıyla ticket senkronizasyon sistemi yapıldı.

TURKCELL METRICUS PROJESİ

Metricus üzerinde işlenen verilerin Performance Point üzerinde ScoreCard'lar aracılığıyla gösterilmesi için manuel olarak ScoreCardlar, Perspective ve KPI'lar birleştirilerek bir Dashboard çıktısı hazırlanıyor. Bu yapıya bir webarayüzüyle Drag-Drop mantığı kullanılarak çok daha kolay bir şekilde Dashboard hazırlanıp PPS Webservice aracılığıyla Performance Point üzerinde oluşturulması.

ABH-COCACOLA PROJESİ

HP Service Manager ile CA Service Desk uygulaması arasında bir ticket entegrasyonu geliştirildi. Bu yapıda CA Service Desk üzerinde ticket açılması için bir Webservice yazıldı. Hp Service Manager üzerinde ticket açılabilmesi için ise HP Service Manager web service kullanıldı.

ECZACIBAŞI – CA SERVICE DESK PROJESİ

Ca Service Desk üzerinde ticket'a eklenen attachmentı webservice kullanarak gönderilmesi ve alınması.

SUPERONLINE – EAM PROJESİ

CA Service Desk üzerine dış kaynaktan Location Insert, Location Update,Ekipman Insert,Ekipman Update yapılabileceği bir Webservice hazırlandı.

ZUGO IMPETUS PROJESİ

Türkiye üzerinde 200 den fazla Renault bayisinin kullandığı müşteri, satış, satış sonrası, cari, e-fatura, raporlama modüllerini içersinde barındıran bir otomasyon uygulamasının geliştirilmesi.

Yetkinlik Bilgileri

C#	:	İleri (Başlangıç / Orta / İleri)
MSSQL	:	İleri (Başlangıç / Orta / İleri)
MYSQL	:	Orta (Başlangıç / Orta / İleri)
ASP.NET,PHP	:	Orta (Başlangıç / Orta / İleri)
SILVERLIGHT-WPF	:	İleri (Başlangıç / Orta / İleri)
JAVA	:	Orta (Başlangıç / Orta / İleri)

Referanslar

Yrd. Doç. Dr. Metin Zontul

İstanbul Aydın Üniversitesi

0530 874 4311