Proje konumuzun başlığını sosyal medya ile pazarlama olarak belirledik. İlk adım sosyal medya ile pazarlamayı araştırmak ve içeriğini anlamak oldu. Yaptığımız araştırmalardan ve sosyal medya ile pazarlamayla uğraşan kişiler ile görüşüldü. Bu alanda işin nasıl yapıldığını ölçümleri neye göre belirledikleri öğrenildi. Bu araştırma bizi konunun bir alt başlığı olan dijital pazarlamaya yöneltti.

Dijital pazarlama çok geniş bir alana hitap ettiği için sosyal medya ile kesiştiği kısımları belirlendi. Biz ise bu kısımlardan okuduğumuz bölüme uygun olan veri madenciliği uygulamalarına yöneldik. Dijital pazarlama ve veri madeniliği ile bilgi çıkarım yöntemlerini araştırdık.

**Burada iş metin madenciliğine döndü.** Dijital pazarlama ile metin madenciliği uygulamalarına baktığımızda tam ve kesin bir olgunluğa ulaşmamış olan sentiment analizi fark ettik. Sentiment analizi seçmemizde ki en büyük etken Türkçe 'nin zorlu bir dil olmasından kaynaklanan sebeplerden ötürü kesin ve %100 çalışan bir yöntemin var olmayışıydı. Bizde bu alanda ki eksikliğe az da olsa bir katkıda bulunabileceğimizi düşündük.

Piyasada var olan sentiment analiz sistemlerine bakıldı. Uygulamaların neyi amaçladığını ve nelere olanak sağladığını öğrendik. Piyasadaki uygulamaların bazıları %80 doğruluk ile duygu analizini gerçekleştirdiklerini söylüyorlar. Ancak İngilizce de dahi %70 'lik doğruluğun çok iyi bir yüzde olduğu kabul edilirken(ingilizce makaleden yola çıkarak söylüyorum) Türkçe gibi sondan eklemeli kelime yapısına sahip, kinaye ve kast etmelerin çok olduğu bir dilde %80lik bir başarı pekte gerçekçi olmadığı kanaatindeyiz.

Böylelikle proje konumuz sosyal medya ile duygu analizine daraltmış olmayı istedik ancak sosyal medya'nın da çok geniş bir alana hitap etmesi, bizi projemiz için ulaşılabilir bir noktaya bir kez daha daraltmaya yöneltti ve tam proje ismi: Sosyal Medya İle Pazarlamada Duygu analizi bir twitter uygulamasına çevrildi.

**Twitter'ı seçmemiz deki etken ise ........**

**TWİT KISMI**

**Proje konumuz net olarak belirlend**i ve sıra uygulamanın geliştirilmesindeydi. İlk olarak analizin gerçekleştirileceği veritabanın oluşturulması gerekiyordu. Twitter'ın kendi bünyesinde yazılımcılar için geliştirdiği bir API servisi olan RestAPI'yi kullanmak istedik. Ancak API'nin az sayıda tweet'e ulaşma izin vermesi çok sayıda tweet için akademik çalışma yapıldığını gösteren evraklarla uğraşmak ve cevap gelmesini beklemek bizim çok vakit alıcıydı. Biz de kendi API sistemimizi yazarak ihtiyacımız olan kadar tweet'e ulaşabilme kararını aldık.

Analiz edilecek verileri elde etme imkanına ulaştıktan sonraki adım ise analizin nasıl gerçekleştirileceği ve puanlamanın nasıl olacağıydı.

Literatür araştırmasına başladık. Konuyla alakalı gerçekleşen eğitimlere elimizden geldiğince katıldık. Metin madenciliğinde sentiment analizi için yapılabilinecek yöntemleri inceledik. Sentiment analizinde sözlük tabanlı analiz kullanmaya karar verdik.

İlk aklımıza gelen fikir TDK'nın sözlüğünden faydalanarak kendimize ait geliştirilebilir bir sözlük oluşturmaktı. Ancak TDK'nın bilgi işlem başkanıyla görüştükten sonra TDK'nın oluşturduğu sözlüğü alamayacağımızı telif hakkı yüzünden pek mümkün olmadığını söyleyerek Türkçe sözlüğe ulaşamayacağımızı bildirdi. (Türkçenin telif hakkının olduğunuda burada öğrendik.) Bizde yönümüzü var olan sözlükler üzerine çevirdik. Daha önce Türkçe de sentiment analiz için oluşturulan sözlükleri inceledik. Balkannet ve Sentiwordnet projelerine ulaştık. Türkçe desteği olan balkannet projesi ilk bakışta sıcak geldi ancak üzerindeki çalışmaların artık durmuş olması ve var olan kelime kapasitesinin yetersiz kalmasından dolayı BalkanNet projesinden faydalanmamaya karar verdik. Balkannet Projesinden vazgeçince ve TDK'nın sözlüğüne ulaşamayınca bizde İngilizce için var olan SentiWordnet projesine göz attık ve bu sözlüğü kullanmaya karar verdik.

SentiWordnet'in en güncel halinin kullanılabilinmesi adına kendimiz "<http://sentiwordnet.isti.cnr.it/>" SentiWordnet için bir API hazırladık bu API İngilizce kelimeyi alıp SentiWordnet'teki puanlarını çekebiliyor duruma geldik.

Ancak senitwordent'teki kelime puanlarına ulaşabilmek ilk önce puanına ulaşmak istediğimiz kelimeyi cümle içinden ayıklamamız ve İngilizceye çevirmemiz gerekiyordu. Ancak kelimelerin imla hatasına sahip olması veya diğer gereksiz olan zamir vs. her cümlede geçen ancak herhangi bir anlam ifade etmeyen kelimelerden temizlenmesi gerekiyordu. Bu temizleme işlemi için var olan, zemberek isimli açık kaynak kodlu ve gönüllüler tarafından geliştirilen programa ulaştık. Zemberek programı ile kurulan cümlede bir imla hatasını gidermekte tam çözüm olmasada, kelimelerin hangileri zamir hangileri sıfat gibi sorularımızı cevaplamamızda ulaşabildiğimiz iyi bir sistemdi. Biz internet üzerinden çalışan dinamik bir sistem yapmak istiyorduk. Sebebi ise zemberek programında veya veritabanında herhangi bir değişiklik olduğu zaman bunu direk sistemimizde görmek istedik bu yüzden online olarak " [http://zemberek-web.appspot.com/](%20http:/zemberek-web.appspot.com/%20)" sitesinden faydalanmak doğru karar olarak görüldü. Bizde zemberek sitesine online olarak ulaşabileceğimiz bir API sistemi geliştirdik.

Kelimelerimiz temizleyip ayırdıktan sonra İngilizceye çevirmemiz gerekiyordu. Bunun içinde ücretsiz bir ingilizce-türkçe çeviri sisteminden ücretsiz ve online olarak faydalandık.

Sıra elde ettiğimiz tweetleri ve bunlara ait kelime puanlamalarını tutabileceğimiz veri tabanını oluşturmaya geldi. Veritabanı için hızlı bir alt yapıya sahip özgürce girdi gönderebileceğimiz bir sisteme ihtiyaç vardı. Bizde NoSQL bir sisteme sahip MongoDB içinde veritabanımızı oluşturmaya karar verdik.

Bütün verileri topladıktan ve gerekli işlemleri gerçekleştirdikten sonra yeterli veri doyumluluğuna ulaşıp, hangi kelimenin hangi twitlerde geçtiğini bulup geçtiği genel olarak geçtiği yerlerde twit ortalamasına göre negatif mi yoksa pozitif mi olduğuna bakılacak ve ortalaması alınarak kendi Veri ambarımızı oluşturacağız. Böylece kendi kendine öğrenme(makine öğrenmesi) gerçekleştiren bir sistemi oluşturmuş olacağız.