**TÜRKİYE AÇIK KAYNAK PLATFORMU**

**ONLİNE YARIŞMA PROGRAMI**

**TÜRKÇE DOĞAL DİL İŞLEME**

**METİN MADENCİLİĞİ LE MEVLANA’NIN MESNEVİ’SİNİN İNCELENMESİ**

**Takım Adı: Teknodil**

**TAKIM ÜYELERİ:**

**LEYLA AKÜZÜM (Takım Lideri)**

**BURAK DEDEOĞLU**

**Leyla AKÜZÜM**

Lisans öğrenimini 2001 yılında Dicle Üniversitesi Eğitim Fakültesinde tamamladı, yüksek lisans öğrenimini de 2021 yılında Dicle Üniversitesi Eğitim Bilimleri Enstitüsünde tamamladı. TUBİTAK 2204b araştırma projeleri kapsamında “Ninelerimiz Anlatıyor” ve “Benim Hayalim: Benim Kitabım” proje isimleriyle yarışmaya katılım göstermiş. Takım üyesinin 2 uluslararası konferanslarda sunulmuş tam metin bildirisi, 3 uluslararası hakemli dergilerde yayımlanmış makalesi ve Uluslararası bir yayınevinde basılmış “Değerli Şiirler” adında bir şiir kitabında editörlüğü bulunmaktadır. Halen Diyarbakır İli Kayapınar Bilim ve Sanat Merkezinde Türkçe öğretmeni olarak görev yapmaktadır.

**Burak DEDEOĞLU**

Lisans öğrenimini 2020 yılında İstanbul Üniversitesi Bilgisayar Mühendisliği alanında tamamladı. Şu an Kale Grubu’nda veri bilimi uzmanı olarak görev yapmaktadır.

**Proje Özeti**

Bu çalışmanın amacı; Türk-İslam dünyasının yetiştirdiği önemli bir şahsiyet olan Mevlana’nın çok derin konuları bile rahat ve anlaşılır bir dille anlattığı Mesnevi’sini veri madenciliği ve metin madenciliği yöntemlerinin farklı yönleri ile ele almaktır. Aynı zamanda dijital çağda dil, edebiyat, tarih ve felsefe alanındaki çalışmalarda da teknolojinin kullanılmasını yaygınlaştırmak ve araştırmacılara bazı öngörüler sağlamak amaçlanmıştır. Proje kapsamında metin madenciliğinin çeşitli yöntemleri kullanılarak metnin söz varlığına ait sınıflamalar, kavramsal analizler ve örüntüler bulunacaktır. Ayrıca edebi metinler sayısallaştırılacak ve Mevlana’nın Mesnevi’si metin verisinden anlamlı bilgiler çıkarmak ihtiyacıyla dil bilim çalışmalarında yazılım yöntemlerinin kullanılmasının önemine dikkat çekilecektir.

**Projenin Odak Aldığı Sorun**

Dil bilimde söz varlığı incelemeleri çeşitli yöntemlerle yapılabilmektedir ancak geleneksel araçlarla büyük miktardaki metin verilerinin depolanması, işlenmesi ve analiz edilmesi emek ve zaman yönünden büyük zahmet gerektirmektedir. Aynı zamanda büyük hacimli verilerden değerli bilgiler elde etmek için uygun modelleri ve eğilimleri belirlemek oldukça zordur (Padhy vd., 2012). Geleneksel veri madenciliği araçları ile bilgi toplamak çok fazla zaman ve çaba gerektirdiğinden bu araçlar metinsel verileri işlemede yetersiz kalmaktadır (Talib vd., 2016, s. 414). Metin madenciliğinin dil bilim araştırmalarında kullanılmasıyla yapılan işlemler, yüksek kalitede yeni bilgi üretimi sağlamaktadır (Anandarajan vd. 2019, s. 2).

**Çözüm**

Geçmişte bir bilginin işlenmesi ve o bilgiden değer üretilmesi çok maliyetli iken günümüzde makine öğrenmesi ve yapay zekâ uygulamaları ile büyük verilerin işlenmesi daha kolay hale gelmiştir. Metin madenciliği, doğal dil metninden anlamlı bilgiler elde etmeye çalışan yeni bir alandır. Metin madenciliği, anahtar kavramları ve temaları yakalamak için veri analizi süreci olarak tanımlamakta ve yazarların bu kavramları ifade etmek için kullandıkları belirli kelimeler veya terimler hakkında önceden bilgi sahibi olmadan gizli ilişkileri ve eğilimleri ortaya çıkarmaktadır (Bose, 2018). Bu bağlamda proje kapsamında metin madenciliğinin çeşitli yöntemlerinin kullanıldığı metnin söz varlığına ait sınıflamalar, kavramsal analizler ve örüntüler yer almaktadır. Ayrıca edebi metinler sayısallaştırılmış ve Mevlana’nın Mesnevi’si metin verisinden anlamlı bilgiler çıkarmak ihtiyacıyla dil bilim çalışmalarında yazılım yöntemlerinin kullanılmasının önemine dikkat çekilmiştir.

Çalışma sonucunda temel olarak ulaşılmak istenen amaçlar aşağıdaki gibidir:

• Edebi metinlerin dil bilimsel açıdan incelenmesi, işlenmesi ve raporlanması çok fazla maliyet ve zaman gerektirir. Bu sebeple edebiyat, tarih ve felsefe gibi disiplinler bakımından kurum ve kuruluşların, araştırmacıların metin madenciliği yöntemleri ile söz konusu çalışmalarına yarar sağlamak.

•Yeni nesil gelişmiş yazılımlar sayesinde büyük metin verilerinin kolayca analiz edilerek metin hakkında genel bilgilere ulaşmak.

• Metin madenciliği yöntemlerini kullanarak yazarın seçtiği kelimeler ve hangi kelimelerin birbiriyle kullandığı gibi karakteristik özellikleri saptamak.

•Metin madenciliğinin hemen hemen her alana uygulanabileceğini göstermek.

• Dijital çağda eğitim, edebiyat, felsefe, ekonomi, siyaset, uluslararası ilişkiler gibi disiplinlere ilişkin güncel olayları açıklamada teknolojinin ve diğer bilimlerin kullanılmasını yaygınlaştırmak.

•Metin madenciliği ve teknikleri ile hem kültürel özelliklerin söz varlığındaki izleri ortaya çıkacak hem de tarihsel sözlüklerin hazırlanması ve geliştirilmesine büyük katkılar sağlayacaktır.

**Yöntem**

Çalışmada emekli Albay Şefik Can tarafından çevirisi yapılan Mevlana’nın Mesnevi’si esas alınmış ve metin sayısallaştırılarak metin işleme yazılımları için hazır duruma getirilmiştir. Eserin tam metni üzerine birçok çalışma bulunmaktadır. Şefik Can tarafından yapılan çeviri metninin bu çalışmada tercih edilmesinin sebebi akademik alanda yaygın biçimde kullanılmasıdır. Metin işleme için farklı yazılımlar kullanılmıştır. Kullanılan veri setinin depolanması için Google firmasına ait bulut tabanlı bir depolama sistemi olan Google Drive programı seçilmiştir. Depolanan veri üzerinde veri analizlerinin yapılması ve gerekli kodların yazılması için Python programlama dili ve bu dile ait NumPy, Pandas, Seaborn gibi veri madenciliğinde sıkça kullanılan kütüphaneler kullanılmıştır. Python programlama dili kullanılarak incelenen metinde bulunan kelimeler, kullanım sıklıkları gibi analizler yapılmıştır.