**Veri Madenciliği Dersi Örnek Proje**

Aşağıda örnek proje konuları, konular ile ilgili makaleler ve veri kümeleri verilmiştir.

|  |  |
| --- | --- |
| **Konu** | **İlgili Makaleler ve Veri Kümeleri** |
| 1. Sahte Kullanıcı Yorumlarının Tespit Edilmesi
 | * Fake Review Detection on Yelp
* Ensemble Classifiers for Spam Review Detection
* <http://odds.cs.stonybrook.edu/yelpchi-dataset/>
 |
| 1. İroni Tespiti
 | * SemEval-2018 Task 3: Irony Detection in English Tweets
* <https://www.kaggle.com/rtatman/ironic-corpus/data>
 |
| 1. Anormallik Tespiti (Sınıflandırma)
 | * Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection
* <https://archive.ics.uci.edu/ml/datasets/detection_of_IoT_botnet_attacks_N_BaIoT>
 |
| 1. Kısa Metinler Üzerinden Konu Çıkarma
 | * Short Text Topic Modeling Techniques, Applications, and Performance: A Survey
* <https://www.kaggle.com/datafiniti/hotel-reviews>
 |
| 1. Sağlık Verilerini Kullanarak Kardiyovasküler Hastalığın Tahmini
 | * Comparison of Machine Learning Models in Prediction of Cardiovascular Disease Using Health Record Data
* <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>
 |
| 1. Film Öneri Sistemi
 | * Content-based movie recommendation system using genre correlation
* <https://grouplens.org/datasets/movielens/>
 |
| 1. Film Başarısının Tahmin Edilmesi
 | * Movie Success Prediction using Data Mining
* <https://www.kaggle.com/rounakbanik/the-movies-dataset>
 |
| 1. Aprori ve Fp Growth Algoritmaları ile Sepet Analizi
 | * Market basket analysis with association rules
* <https://www.kaggle.com/irfanasrullah/groceries>
 |
| 1. Veri Madenciliği Yöntemleri ile Satış Tahmini Gerçekleştirme
 | * A Comparative Study of Univariate Time-series Methods for Sales Forecasting
* <https://www.kaggle.com/tevecsystems/retail-sales-forecasting>
 |
| 1. Pazarlama Stratejisi Geliştirmek için Müşterilerin Kümelenmesi (EM (Expectation-Maximization), Bulanık C-Kümeleme, DB-Scan)
 | * Customer clustering for a new method of marketing strategy support within the courier business
* <https://www.kaggle.com/arjunbhasin2013/ccdata>
 |
| 1. Diyabet Veri Kümesi Üzerinden Sınıflandırma Yapılması
 | * A Parallel Ensemble Fuzzy Classifier for Diabetes Diagnosis
* <https://www.kaggle.com/uciml/pima-indians-diabetes-database>
 |
| 1. Yağış Verileri Üzerinden Tahmin Yapılması
 | * Rainfall Forecasting Through ANN and SVM in Bolangir Watershed, India
* <https://www.kaggle.com/redikod/historical-rainfall-data-in-bangladesh>
 |
| 1. Sahtecilik Tespiti
 | * Analysis of Techniques for Credit Card Fraud Detection: A Data Mining Perspective
* <https://www.kaggle.com/mlg-ulb/creditcardfraud>
 |
| 1. Topluluk Tespiti
 | * Community detection in graphs
* <https://www.kaggle.com/ahajib/travian-dataset>
 |