|  |
| --- |
| **Ad-Soyad**: Hakan Öznam |
| **Proje Başlığı**: Makine Öğrenmesi ile Satranç Oyun Sonu Tahmini |

PROJE 1

**İçindekiler**

[**Projenin Konusu ve Amacı** 2](#_Toc54594196)

[**Projede Kullanılan Materyaller** 3](#_Toc54594197)

[Veri Kümesi ve Özellikleri 3](#_Toc54594198)

[**Proje Geliştirme Aşamaları** 4](#_Toc54594200)

[**Proje ile Ulaşılan Sonuçlar** 5](#_Toc54594201)

# **Projenin Konusu ve Amacı**

Bu projede satranç oyunlarının analizi yapılacaktır. [Lichess](https://lichess.org/) adlı satranç sitesinde oynanan 20000’e yakın sayıda maç veri kümesi olarak ele alınmıştır. Veri kümesi [Kaggle](https://www.kaggle.com/datasnaek/chess) aracılığıyla elde edilmiştir.

Tek bir satranç oyununda birden fazla oyundan oluşan bir veri kümesinin dışında birçok bilgi de bulunur. Projenin amacı ise bu verileri analize uygun bir şekilde hazırlamaktır. Örneğin;

* Bir oyuncunun siyah veya beyaz olmasının kazanmasına olan etkisi
* En çok hangi oyun açılışının tercih edilmesi
* Oyun açılışları ile zafer arasındaki ilişki
* Yüksek reytingli oyuncuların maçlarının hamle sayılarına etkisi
* Berabere biten maçların oyun analizleri

Bu projede gösterilen sonuçlara ulaşabilmek için veri kümesi formata uygun düzenlenmesi hedeflenmektedir.

# **Projede Kullanılan Materyaller**

## Veri Kümesi ve Özellikleri

Veri kümesi 20058 satır ve 13 sütundan oluşmaktadır. Veri kümesinin sahip olduğu özellikler ise;

**Id\*** 🡪 Karşılaşmanın tekrarının Lichess sitesinde bulunmasını sağlayan kod

**rated** 🡪 Karşılaşmada reyting kazanımı olacak mı? (Evet/Hayır)

**game\_time** 🡪 Karşılaşma süresi

**turns** 🡪 Oyunda geçen hamle sayısı

**victory\_status** 🡪 Oyun sonucu

**winner** 🡪 Kazanan taraf veya varsa beraberlik

**white\_id\*** 🡪 Beyaz oyuncu kimliği

**white\_rating** 🡪 Beyaz oyuncu reytingi

**black\_id\*** 🡪 Siyah oyuncu kimliği

**black\_rating** 🡪 Siyah oyuncu reytingi

**moves\*** 🡪 Karşılaşmadaki tüm hareketler

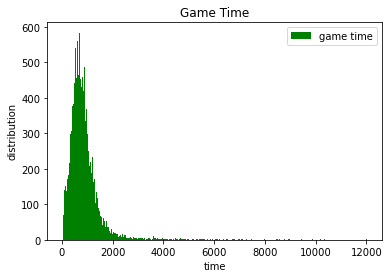
**opening\_eco** 🡪 Oyun açılışlarının satranç ansiklopedisine göre kodları

**opening\_name** 🡪 Oyun açılışlarının adı

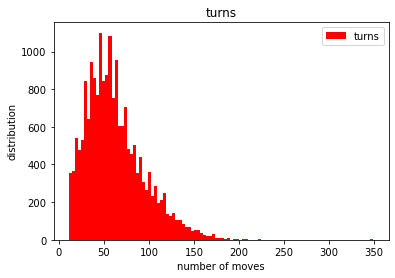
(\*) ile işaretli olan özellikler analiz kısmında kullanılmayacaktır. O özellikler internet sitesinden maçlara ulaşabilmek için kullanılacaktır.

Bazı özelliklerin grafiksel gösterimi şu şekildedir;

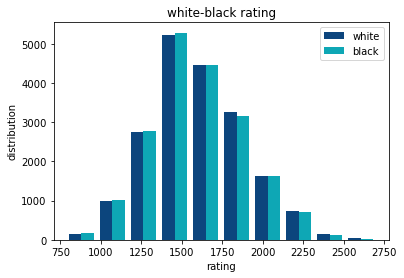
* *Oyun sürelerinin genel dağılımı*



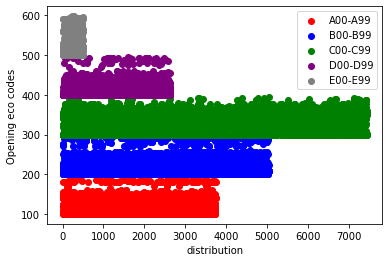
* *Hamle sayılarının genel dağılımı*



* *Beyaz ve siyah oyuncuların reytingleri*



* *Oyun açılışlarını belirten eco kodlarının kullanım sıklığı*



**Proje Geliştirme Aşamaları**

1. **Aşama(Veri Temizleme)**

Veri kümesi Excel’den alındıktan sonra analiz sonuçlarının tutarlı olması için hamle sayısı 10’dan küçük olan maçları veri kümesinden budama işlemi yapılır. Bu adımda 681 satır veri silme işlemi yapıldı.

1. **Aşama(Eksik Veri İşleme)**

Eksik verileri tespit etmek için yapılan işlem sayesinde game\_time(oyun süresi) sütununda 9000 tane verinin eksik olduğu görülür.

Hamle sayısı arttıkça oyun sürelerinin arttığı gözlemlenir. Bu nedenle veri kümesindeki eksik verileri doldurmak için diğer maçlardaki hamle sayısı ile oyun süresi arasındaki ilişki incelenir. Eksik verileri doldurmak için IterativeImputer metodu kullanılır. Metot bir nevi tahminleyici görevi görür. Bu metot oyunlardaki hamle sayısı ile oyun süresi arasındaki ilişkiyi analiz eder. Daha sonra elde ettiği bu analiz sonucuna göre eksik verileri doldurur. Bu aşamada eksik verilerin doldurulması işlemi gerçekleştirilmiştir.

1. **Aşama(Kategorik Veri Dönüştürme)**

Veri kümesinde analize dahil edilecek 4 özellik kategorik veri tiplerine sahiptir. Bu özellikleri sayısal veri tiplerine dönüştürme işlemi yapılacaktır.

Rated ve winner özellikleri için LabelEncoder metodu kullanılarak sayısal verilere dönüştürme işlemleri yapılır. LabelEncoder metodu veriyi birebir sayısallaştırmaya yarar. Uygulandığı sütundaki her bir farklı veriye başka bir sayısal değer ataması gerçekleştirir.

Rated özelliği için (False 0; True 1) değerlerini almıştır.

Winner özelliği için (Black 0; Draw 1; White 2) değerlerini almıştır.

Victory\_status özelliği 4 farklı içeriğe sahiptir. Bu özellik için OneHotEncoder metodu kullanılır. OneHotEncoder metodu mevcut değere “1” verip mevcut olmayanlara ise “0” değeri vermektedir.

outoftime => [0,0,1,0]; resign => [0,0,0,1]

mate => [0,1,0,0]; draw => [1,0,0,0] değerlerine sahip olur.

Opening\_eco özelliği 365 farklı içeriğe sahiptir. Özelliğin değerleri (bir kategorik veri ve iki nümerik veri olmak üzere) 3 haneden oluşmaktadır. Örneğin; “A00”,”B12”,”C26”,”D50”,”E36”. Bu özellik için ne LabelEncoder ne de OneHotEncoder kullanımı uygundur. Bu yüzden başka bir metot kullanılmaya karar verilmiştir. Verilerde A yerine 1, B yerine 2, C yerine 3, D yerine 4, E yerine 5 değeri koyulmuştur. Bu şekilde kategorik veriler nümerik hale dönüştürülmüştür.

1. **Aşama(Verilerin Birleştirilmesi)**

İlk üç aşamadaki verileri işleyebilmek için veriler dağıtıldı. Formata uygun hale getirilebilmesi için veriler işlendi. Bu aşamada ise dağıtılan verilerin bir araya getirilerek analiz edilecek veri kümesi elde edildi. Bu veri kümesi 8 özellik ve 19377 satırdan oluşmaktadır.

# **Proje ile Ulaşılan Sonuçlar**

Analiz yapılacak veri kümesi bulundu. Analizin doğru sonuç verebilmesi için bazı veriler silindi. Daha sonra eksik veriler formata uygun bir şekilde işlendi. Kategorik veriler sayısal verilere dönüştürüldü. En son veriler birleştirilerek analize uygun bir hale getirildi.