

VERİ MADENCİLİĞİNE GİRİŞ DERSİ PROJE ÖDEV RAPORU

Karar Ağaçları Metodu:

Beyaz Şah ve Kaleye Karşı Siyah Şah Satranç Oyun Sonu

Rumeysa Emine Şahin – 19360859002

Bursa Teknik Üniversitesi

Özet

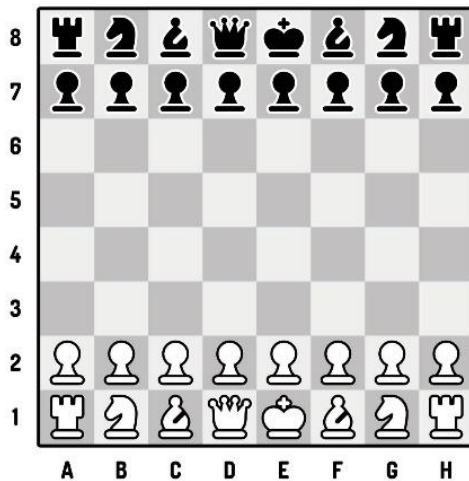
Oyunsonu veritabanları, taşların tahtadaki konumları için depolanmış oyun-teorik değerlerin tablolarıdır. Kaydedilen oyun teorisi değerleri, konumların her iki taraf için de kazanılıp kazanılmadığını gösterir veya minimum-maksimum-optimal oyunu varsayarak kazanma derinliğini (hamle sayısını) içerir.

Satranç son oyunu, tahtada birkaç taşın kaldığı bir durumdur. Karşı tarafın Şah-Kale'sine karşı Şah'a sahip olmak yaygın bir satranç oyunsonudur.

Satranç oyununun sonucunu tahmin etmek için farklı algoritmalar kullanan birçok çalışma vardır. Bu çalışma Satranç Oyunsonu probleminin karar ağaçları ile sınıflandırılması üzerinedir.

Satranç Oyunu

Satranç, iki oyuncu arasında satranç tahtası ve taşları ile oynanan bir masa oyunudur. Bu oyun satranç tahtası denilen 8x8'lik kare bir alan üzerinde oynanır. Satranç tahtası sekiz sütun (a,b,c,d,e,f,g,h) ve sekiz sıra (1,2,3,4,5,6,7,8) içerir. Toplam 64 karenin yarısı siyah, yarısı beyaz renklerden oluşur. 32 adet satranç taşıyla oynanır. Taraflar beyaz ve siyah renkli taşları alır, her oyuncunun bir seferde bir hamle yapmasıyla oyun gelişir. Oyunun başında beyaz ve siyahların 16 taşı bulunur. Bunlar bir şah, bir vezir, iki kale, iki fil, iki at ve sekiz piyondan oluşur. Satranç Oyununda her oyuncunun hedefi, rakibin şahını tuzağa düşürerek bir mat yapmaktır. Yaygın bir satranç oyunsonu durumu, (Şah-Kale)'ye karşı (Şah)'a sahip olmaktır. (1)



Karar Ağaçları

Karar ağaçları, makine öğrenmesinde sınıflandırma ve regresyon problemlerini çözmek için kullanılan popüler bir öğrenme algoritmasıdır. Karar ağaçları, bir dizi karar kuralı ve veri özelliklerini kullanarak bir veri noktasını sınıflandırmak veya bir değeri tahmin etmek için bir ağaç yapısı oluşturur. Karar ağacı oluşturma işlemi tamamlandığında, ağaç veri kümesindeki örneklerin sınıflandırılması veya değer tahmini için kullanılır. Bir veri noktası ağaç boyunca ilerlerken, her düğümdeki karar kuralı kullanılır ve uygun alt düğüme gidilir. Bu şekilde, son düğüme ulaşılır ve sonuç elde edilir. Karar ağaçları, sınıflandırma problemlerinde sınıf etiketlerini tahmin etmek için kullanılabilir.

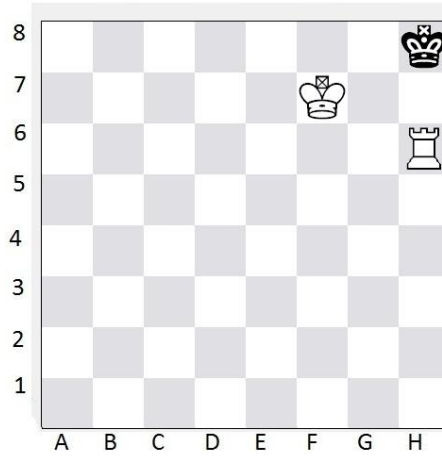
Veri Seti Bilgileri

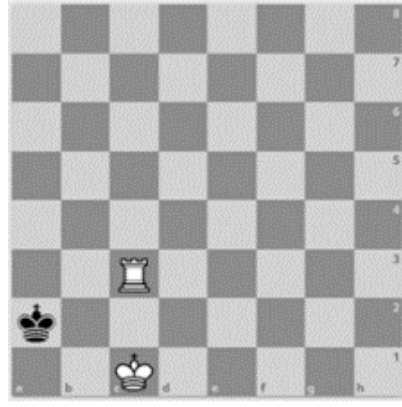
- Veri seti Turing Enstitüsü, Glasgow, İngiltere'de Michael Bain ve Arthur van Hoff tarafından oluşturulmuştur. Veri setinin kaynağı, UCI Machine Learning Repository'de bulunmaktadır. (2)
- Tarih: Haziran 1994
- Örnek Sayısı: 28056
- Altı öznitelik değişkeni ve bir sınıf değişkeni vardır.
- Eksik öznitelik değerleri yoktur.

Öznitelik Bilgileri

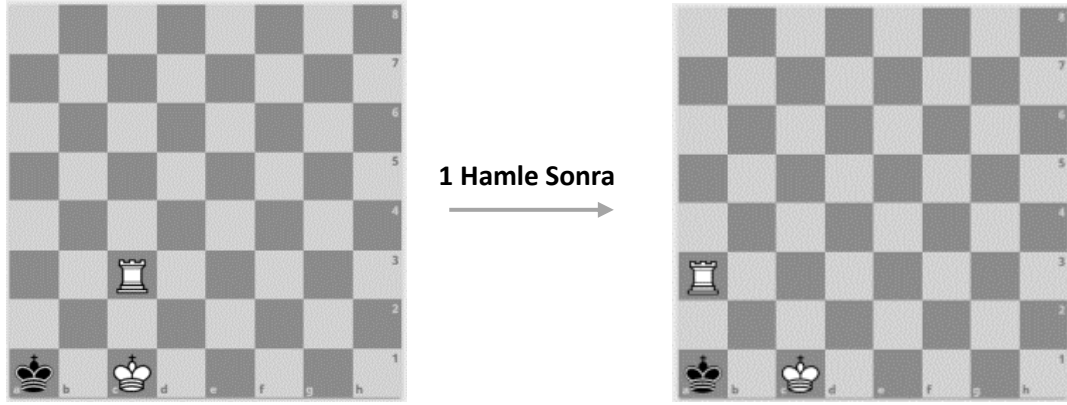
1. Beyaz Kral Sütun
2. Beyaz Kral Satır
3. Beyaz Kale Sütun
4. Beyaz Kale Satır
5. Siyah Kral Sütun
6. Siyah Kral Satır

Tahtadaki her bir taşın konumu, sütun ve satır kullanılarak belirtilir. Bu veri seti için 1. ve 2. öznitelik değerleri beyaz kralın tahtadaki konumunu, 3. ve 4. öznitelik değerleri beyaz kalenin konumunu, 5. ve 6. öznitelik değerleri ise siyah kralın konumunu belirtir.





Yukarıdaki santraç oyun sonu için verisetini inceleyelim. Şah herhangi bir yönde (yatay, dikey veya çapraz) bir kare hareket edebilirken, Kale dikey veya yatay yönde herhangi bir sayıda kare hareket edebilir. Veri Kümesine göre bir sonraki hamleyi oynayacak olan oyuncu Siyah Şah'ı kontrol edecektir.

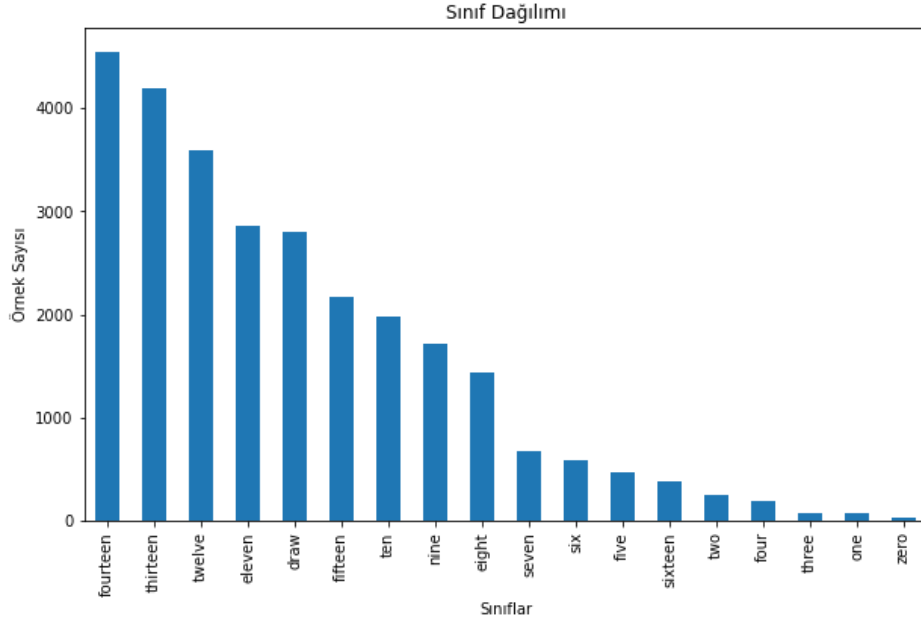


Siyah şah Beyaz Kale yüzünden A3, B3'e ve Beyaz Şah yüzünden B1, B2'ye hareket edemez. Sadece A1'e hareket edebilir. Bu durumda, Beyaz Kale C3'ten A3'e hareket eder ve beyaz oyunu tek hamlede kazanır. (3)

Bu çalışma aşağıdaki adımlardan oluşmaktadır:

- 1.Adım : Veri Setini İndirme Ve İnceleme
- 2.Adım : Veri Setinin Keşfi Ve Analizi
- 3.Adım : Veri Setinin Görselleştirilmesi
- 4.Adım : Veri Ön işleme
- 5.Adım : Veri Setinin Bölünmesi
- 6.Adım : Karar Ağacı Modelini Oluşturma ve Eğitim
- 7.Adım : Karar Ağacı Modelinin Değerlendirilmesi

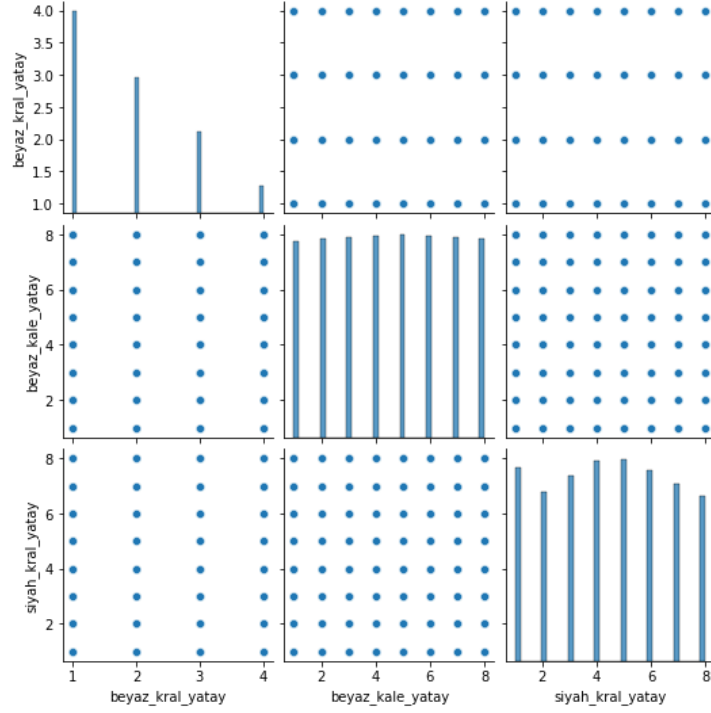
Veri Setindeki Sınıf Dağılımı:



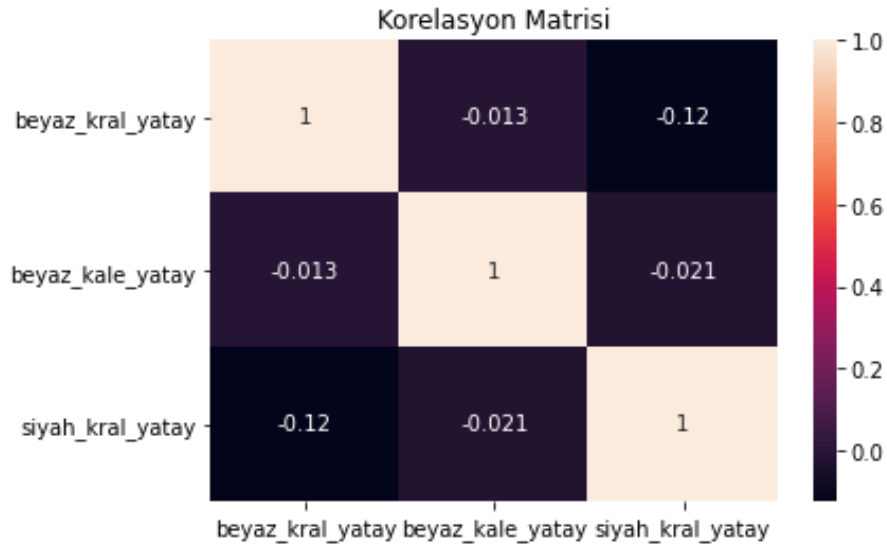
Yukarıdaki grafik veri setindeki her bir sınıfın kaç örnek ile temsil edildiğini gösterir. Beyazın 14 hamlede kazandığı oyun sonu senaryolarının daha fazla olduğunu görebiliriz.

Veri Setindeki Özniteliklerin Birbiriyle ilişkisi:

Aşağıdaki grafik özniteliklerin dağılımlarını ve birbirleriyle ilişkilerini gösterir.



Korelasyon matrisi, veri setindeki öznitelikler arasındaki ilişkileri gösteren bir matristir. Korelasyon katsayıları, her öznitelik çifti için bir değer içerir. Değerler genellikle -1 ile +1 arasında değişir. Pozitif bir değer, iki öznenitenliğin pozitif bir ilişkiye sahip olduğunu, negatif bir değer ise negatif bir ilişkiye sahip olduğunu gösterir. 0 değeri, iki öznenitenlik arasında bir ilişkinin olmadığını gösterir.



Sonuç

Karar ağacı modelinin değerlendirilmesi sonucu aşağıdaki sonuçlar elde edilmiştir:

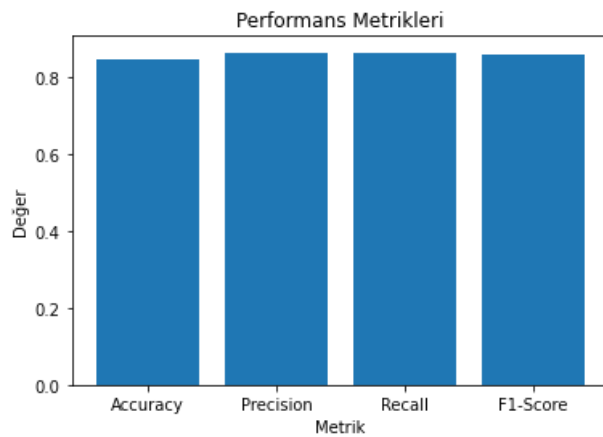
Accuracy: 0.848

Precision: 0.863

Recall: 0.865

F1 Score: 0.86

Sonuçlara bakarak, modelimizin genel olarak iyi bir performans sergilediğini ve verileri doğru şekilde sınıflandırabildiğini söyleyebiliriz.



Modelin sınıflandırma sonuçlarını gerçek değerlerle karşılaştırmak için karmaşıklık matrisi:

Karmaşıklık Matrisi:

```
[ [403 11 22 26 2 0 17 10 0 8 4 3 13 17 0 12 1 0]
[ 10 250 1 0 0 0 0 10 0 12 1 0 2 1 0 0 0 0]
[ 13 2 507 1 0 0 3 10 0 0 0 0 19 12 0 42 0 0]
[ 10 0 1 358 0 0 28 0 0 0 0 8 0 1 0 0 0 0]
[ 0 1 0 0 87 0 0 0 1 0 1 0 0 0 2 0 0 0]
[ 2 0 0 0 3 31 0 0 0 0 1 0 1 0 3 0 0 0]
[ 8 0 0 39 0 0 822 0 0 0 0 2 45 0 6 0 0 0]
[ 6 12 7 0 0 0 0 295 0 1 2 0 19 0 0 4 0 0]
[ 0 1 0 0 0 0 0 0 20 0 0 0 0 0 0 0 1 0]
[ 4 9 0 0 2 0 0 3 0 106 4 0 0 0 0 1 0 0]
[ 2 2 0 0 1 1 0 0 0 6 116 0 0 0 0 0 0 0]
[ 2 0 0 7 0 0 0 0 0 0 0 67 0 0 0 0 0 0]
[ 6 2 26 0 1 0 0 16 0 1 2 0 314 0 0 3 0 0]
[ 7 0 10 4 1 0 58 4 0 0 0 0 0 688 0 49 0 0]
[ 0 0 0 0 0 0 0 0 0 0 0 0 0 0 12 0 0 0]
[ 15 2 34 0 0 0 11 2 0 0 0 0 7 40 0 635 0 0]
[ 2 0 0 0 0 0 0 0 0 0 0 0 0 0 2 0 45 0]
[ 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 5]]
```

Aynı veri seti üzerinde (King Rook vs King) bir grup algoritma uygulanarak çok sınıflı sınıflandırıcıların doğruluğunu test eden başka bir akademik çalışmanın sonuçları ise aşağıdaki gibidir : (3)

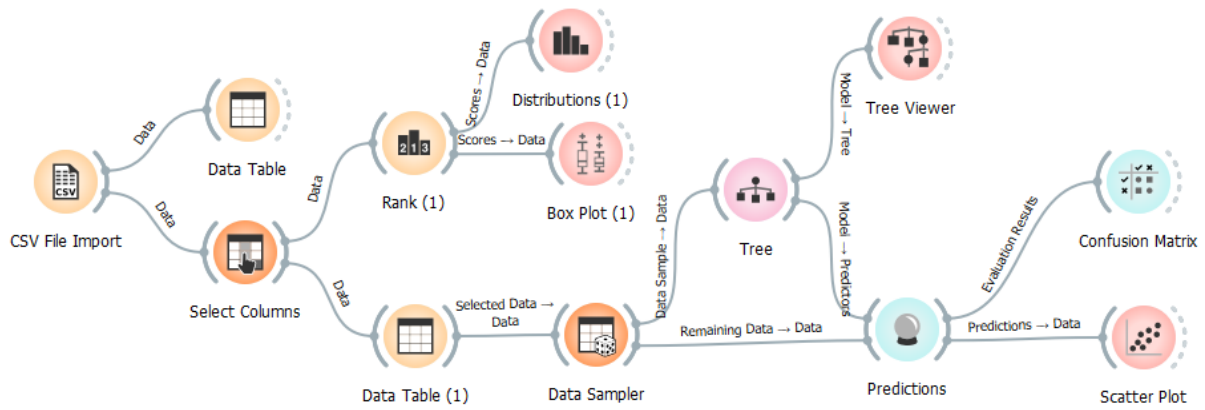
Algorithm	Overall Accuracy	Average Accuracy
Multiclass Decision Jungle	0.496376	0.944042
Multiclass Decision Forest	0.793038	0.977004
Logistic Regression	0.321255	0.924584
Neural Network	0.622668	0.958074

Yukarıdaki grafiğe göre diğer algoritmalar ile karşılaştırıldığında en iyi doğruluğu sağlayan algoritma karar ağaçlarıdır.

Yukarıda çalışmanın karar ağaçları metoduna ait sonuçlarının bu çalışmada elde edilen sonuçlar ile karşılaştırılması aşağıdaki grafikte verilmiştir.

Değerlendirme Ölçütü	Geçerli Çalışma	Diğer Akademik Çalışma
Accuracy	0.848	0.977
Precision	0.863	0.799
Recall	0.865	0.793

Orange Programı Kullanılarak Elde Edilen Sonuçlar



		#	Info. gain	Gain ratio	Gini	χ^2
1	C X.0	4	0.165	0.094	0.012	2134.415
2	N X.1		0.290	0.167	0.033	7640.705
3	C X.2	8	0.050	0.017	0.005	606.480
4	N X.3		0.021	0.010	0.002	195.059
5	C X.4	8	0.185	0.063	0.016	2124.714
6	N X.5		0.270	0.135	0.022	4733.162

Model	AUC	CA	F1	Precision	Recall
Tree	0.924	0.584	0.581	0.589	0.584

Confusion matrix for Tree (showing number of instances)

		Predicted																		
		draw	eight	eleven	fifteen	five	four	fourteen	nine	one	seven	six	sixteen	ten	thirteen	three	twelve	two	zero	Σ
Actual	draw	288	15	34	62	3	3	121	15	1	8	8	9	11	51	0	65	0	2	696
	eight	30	254	19	0	3	0	1	40	0	2	3	0	4	3	0	15	0	0	374
	eleven	32	6	350	2	0	1	24	36	2	0	0	0	35	54	0	169	0	0	711
	fifteen	28	0	1	375	0	0	127	0	0	0	0	7	0	8	0	3	0	0	549
	five	3	0	0	0	88	6	0	0	0	2	9	0	0	2	0	4	0	0	114
	four	4	1	2	0	22	13	0	0	0	1	5	0	0	0	1	1	0	0	50
	fourteen	39	1	3	114	0	0	869	1	0	0	0	1	5	98	0	42	0	0	1173
	nine	32	22	40	0	1	1	3	248	0	9	1	0	20	12	0	21	0	0	410
	one	0	0	2	0	1	0	0	0	15	0	0	0	0	0	0	0	3	0	21
	seven	24	29	21	0	5	1	2	6	0	81	7	0	5	0	0	5	0	0	186
	six	16	4	3	0	17	0	0	0	0	8	81	0	3	2	0	0	1	0	135
	sixteen	8	0	0	38	0	0	3	0	0	0	0	55	0	1	0	0	0	0	105
	ten	43	11	82	0	1	0	10	44	0	4	0	0	209	11	0	55	0	0	470
	thirteen	58	3	20	15	0	0	213	5	0	0	0	0	9	572	0	176	0	0	1071
	three	1	0	0	0	2	4	0	0	0	1	0	0	0	0	7	0	2	0	17
	twelve	42	1	78	9	1	0	44	14	0	0	1	0	19	127	0	539	0	0	875
	two	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	51	0	53
	zero	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	4
	Σ	648	347	655	615	144	29	1417	409	18	116	115	72	320	941	10	1095	57	6	7014

Kaynakça

1. <https://tr.wikipedia.org/wiki/Satran%C3%A7>. [Çevrimiçi]
2. <https://archive.ics.uci.edu/ml/datasets/Chess+%28King-Rook+vs.+King%29>. [Çevrimiçi]
3. <https://arxiv.org/ftp/arxiv/papers/2111/2111.05976.pdf>. [Çevrimiçi]
4. <https://www.youtube.com/playlist?list=PLmNPvQr9Tf-ZSDLwOzxpY-HrE0yv-8Fy>