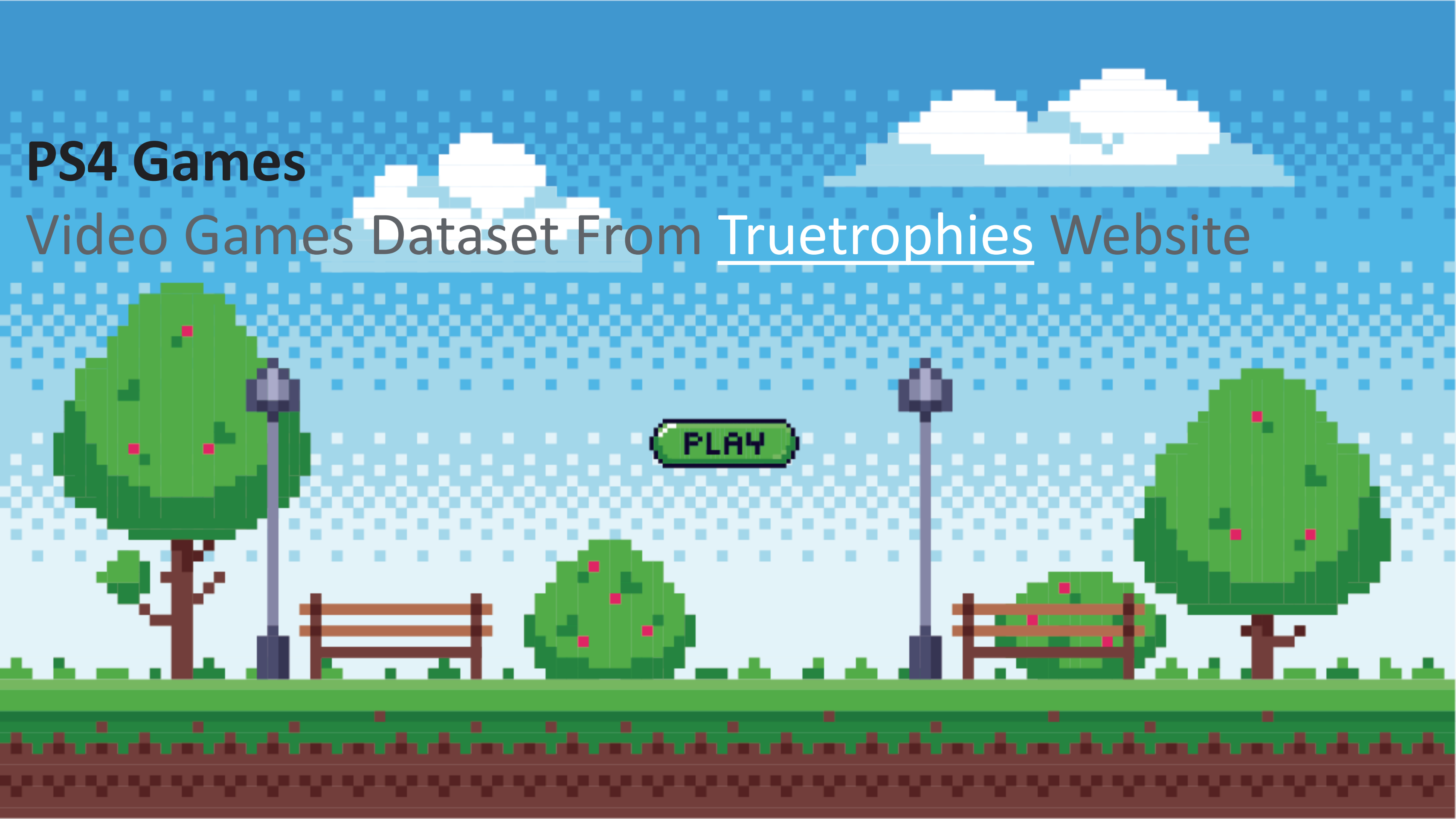


İSTANBUL TEKNİK ÜNİVERSİTESİ
BÜYÜK VERİ & İŞ ANALİTİĞİ UZMANLIĞI
SERTİFİKA PROGRAMI
PROJE SUNUMU

KUAR DALİDA DİKİCİ

69. DÖNEM



PS4 Games

Video Games Dataset From Truetrophies Website

Adımlar/ İçindekiler

1. Veri seti seçildi ve indirildi.

(Yüksek Lisans Tez Konum ile bağlantılı olarak oyunlar üzerine bir veri seti tercih edildi.)

<https://www.kaggle.com/datasets/ww1234/ps4-games>

2. Verilerle ilgili genel bilgiler özetlendi.

3. İndirilen veriler müşteri analitiği odağında **KNIME Analytics** kullanılarak incelendi.

4. Veri seti betimleyici adımlar kullanılarak tablo ve grafiklerle tanıtıldı.

5. Veri setinde eksik verilerin varlığı kontrol edildi.

6. Verilerin anlamlandırılması, uygun problemin ne amaçla seçildiği üzerine:

«Oyunlar» ve «Oyuncu tipleri» ile ilgili literatürden bilgiler eklendi.

7. Veri setine uygun problem belirlendi.

8. Probleme uygun 3 model belirlendi.

I. Lineer Regresyon / «**Rayting**» üzerine en iyi fonksiyonu ve değişkenleri tespit etme.

II.Karar Ağacı (Forward Feature Selection ile iyileştirme) / **Achiever'lar için oyun önerisi: Sınıflandırma**

III.Logistic Regresyon (K-Fold Validation) / **Achiever'lar için oyun önerisi: Sınıflandırma**

9. Modellerin performansları metriklere göre karşılaştırıldı.

I. MSE

II.MAE

III.R²

IV.Accuracy

10.İş Analitiği bağlamında öneriler listelendi.

1.Data Set seçildi ve indirildi.

<https://www.kaggle.com/datasets/ww1234/ps4-games>

2.Verilerle ilgili genel bilgiler özetlendi.



Veri Seti Hakkında

Bağlam

- Bu veri kümesi, şu an için PlayStation 4 konsoluna dahil tüm oyunları içerir.
- Bu veri setini oluşturmak için [trueachievements](http://www.trueachievements.com) web sitesinden faydalanılmıştır.

İçerik

Tablodaki sütunların içeriği ise şöyle:

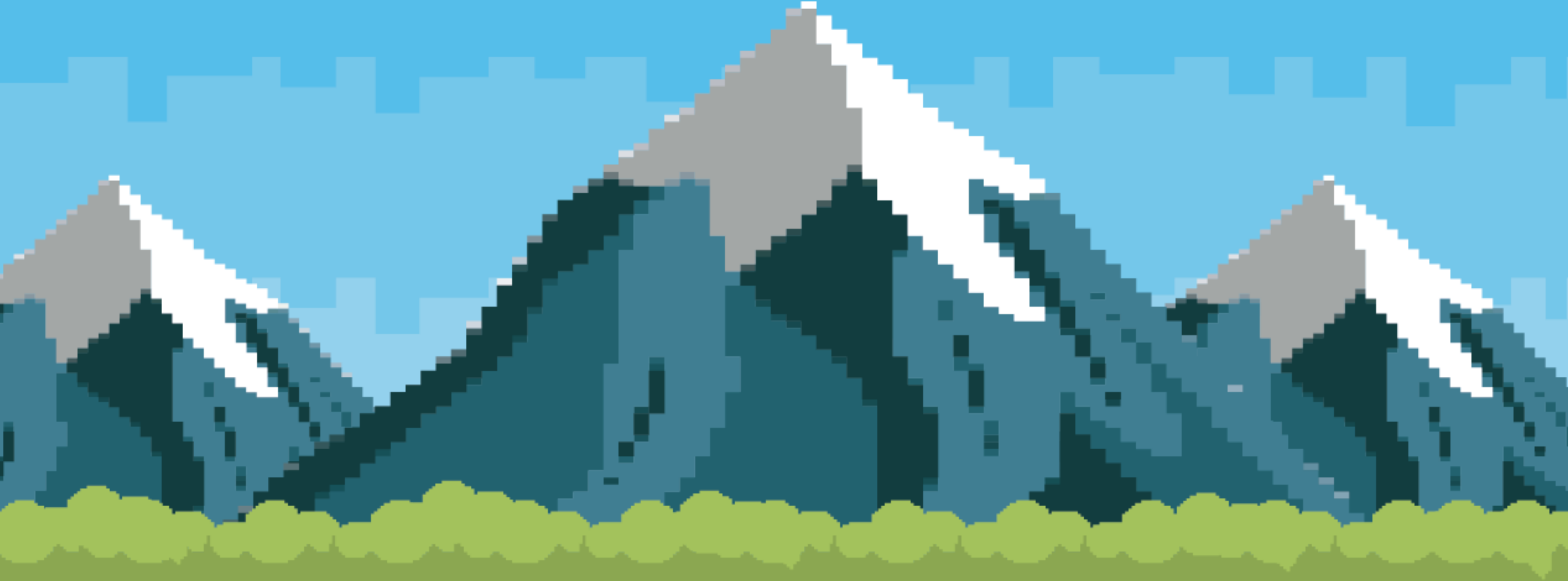
Table "default" - Rows: 1584 Spec - Columns: 9			Property
Columns: 9	Column Type	Column Index	Açıklama
game	String	0 →	PS4 Güncel Oyun isimleri
score	Number (integer)	1 →	İlgili oyunda elde edilebilecek maksimum ganimet* sayısı
leaderbord	Number (double)	2 →	Oyunda maksimum score'u elde etmişlerin liderlik tablosu
gamers	Number (integer)	3 →	İlgili oyunda Truetrophy websitesine kayıtlı oyuncu sayısı
comp_perc	Number (double)	4 →	İlgili oyunu tamamlayanların yüzdesi
rating	Number (double)	5 →	İlgili oyunun reytingi (popülerliği)
url	String	6 →	İlgili oyunu yakından incelemek için URL
min_comp_time	Number (integer)	7 →	İlgili oyunda tüm maksimum score'u elde etmek için geçirilmesi gereken min. saat
max_comp_time	Number (integer)	8 →	İlgili oyunda tüm maksimum score'u elde etmek için geçirilmesi gereken max. saat

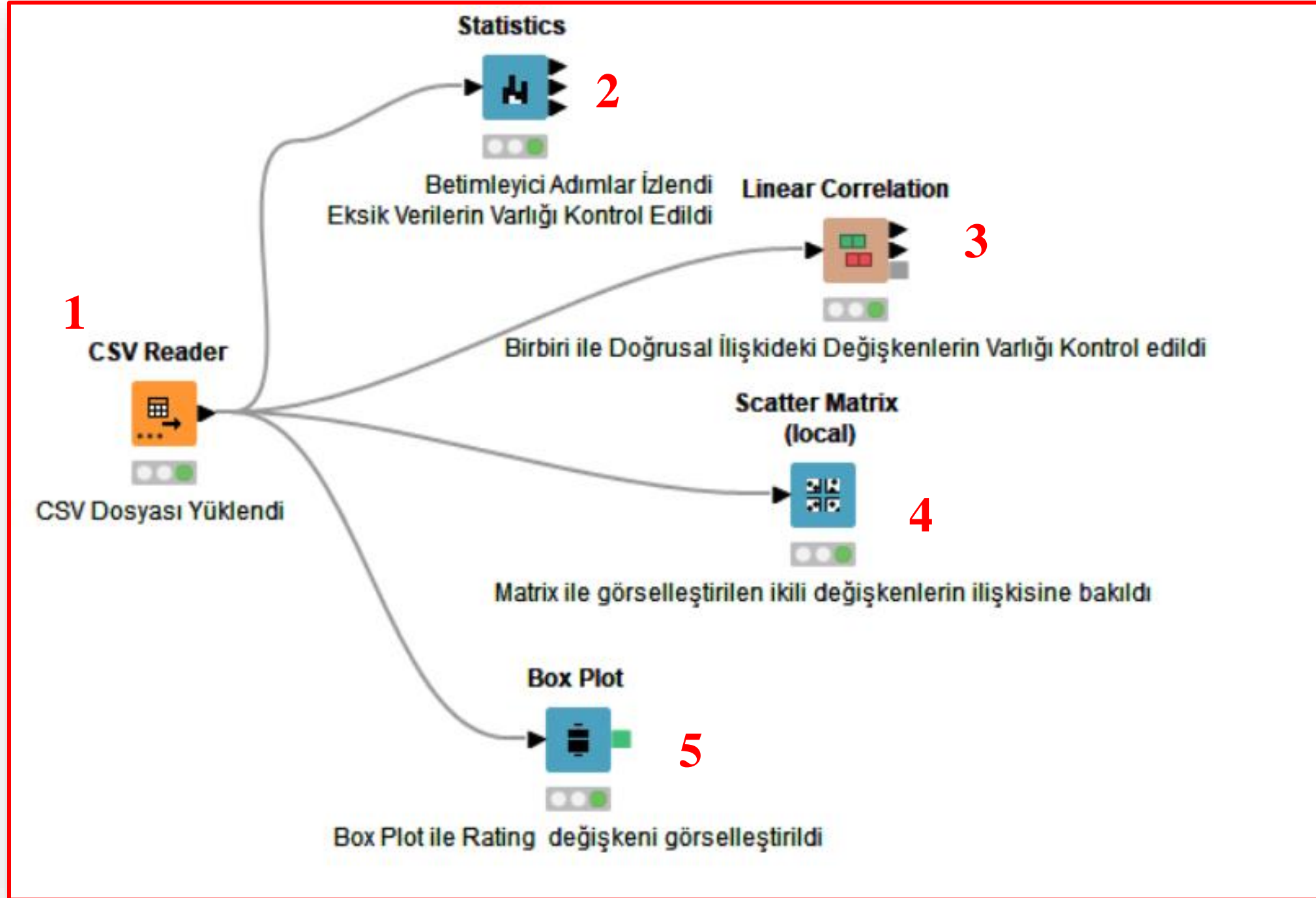
GANİMETLER_TROPHIES

Aşağıda gördüğünüz objeler çeşitli oyunlarda görebileceğiniz ganimetlerdir. Bu ganimetleri topladıkça oyunda bir anda bir ses duyarak çeşitli başarılar elde edebilirsiniz. Oyuncu tiplerinden «Achievers» bu toplama ve ses için motive olan ve oyunda kalan oyunculardır. Verilerin analizinde de problemi belirlerken bu perspektiften yola çıkıldı.



3. İndirilen veriler müşteri analitiği odağında KNIME Analytics kullanılarak incelendi.
4. Veri seti betimleyici adımlar kullanılarak tablo ve grafiklerle tanıtıldı.
5. Veri setinde eksik verilerin varlığı kontrol edildi.





Rakamlarla işaretlenmiş bölümlerle (Node'larla) ilgili açıklamalar sonraki yansılarda detaylandırılacaktır.

- 1 CSV formatında indirilen veriler tablo halinde yüklenmiştir. Bu veriler Kaggle sitesinden indirilmiştir. Veriler Truetrophy websitesinden alınmıştır. Toplamda 1584 satır ve 9 sütundan oluşmaktadır. Eksik veri bulunmamaktadır.

File Table - 3:1 - CSV Reader

File Edit Hilite Navigation View

Table "default" - Rows: 1584 Spec - Columns: 9 Properties Flow Variables

Row ID	S game	I score	D leaderb...	I gamers	D comp_p...	D rating	S url	I min_co...	I max_co...
0	A Boy and His Blob	638	2.02	2194	16.5	3.2	https://www.truetrophies.com/game/A-Boy-and-His-Blob/trophies	15	20
1	A Hat in Time	1992	1.53	7062	35.9	4.2	https://www.truetrophies.com/game/A-Hat-in-Time/trophies	15	20
2	A Hero and a Ga...	1364	1.01	503	97.6	5	https://www.truetrophies.com/game/A-Hero-and-a-Garden/tro...	0	1
3	A Hero and a Ga...	1363	1.01	581	97.8	2.9	https://www.truetrophies.com/game/A-Hero-and-a-Garden-Eur...	0	1
4	A King's Tale: Fin...	637	2.02	21914	14.1	3.3	https://www.truetrophies.com/game/A-Kings-Tale-Final-Fantas...	4	5
5	A Plague Tale: I...	1691	1.34	22221	44.3	4.3	https://www.truetrophies.com/game/A-Plague-Tale-Innocence/...	15	20
6	A Summer with t...	1372	1.02	640	96.1	2.7	https://www.truetrophies.com/game/A-Summer-with-the-Shiba-...	0	1
7	A Summer with t...	1359	1.01	377	98.7	2.8	https://www.truetrophies.com/game/A-Summer-with-the-Shiba-...	0	1
8	A Summer with t...	1376	1.02	719	95.7	1.7	https://www.truetrophies.com/game/A-Summer-with-the-Shiba-...	0	1
9	A Way Out	1793	1.36	51657	46.6	4	https://www.truetrophies.com/game/A-Way-Out/trophies	6	8
10	A Winter's Daydr...	1303	1.01	2249	98	2.1	https://www.truetrophies.com/game/A-Winters-Daydream/trop...	0	1

Table "default" - Rows: 1584 Spec - Columns: 9 Properties Flow Variables

Columns: 9	Column Type	Column Index	Color Handler	Size Handler	Shape Han...	Filter Handler	Lower Bound	Upper Bound
game	String	0					?	?
score	Number (integer)	1					189	12,017
leaderbord	Number (double)	2					1	7.38
gamers	Number (integer)	3					60	245,220
comp_perc	Number (double)	4					0	99.6
rating	Number (double)	5					0.5	5
url	String	6					?	?
min_comp_time	Number (integer)	7					0	200
max_comp_time	Number (integer)	8					1	700

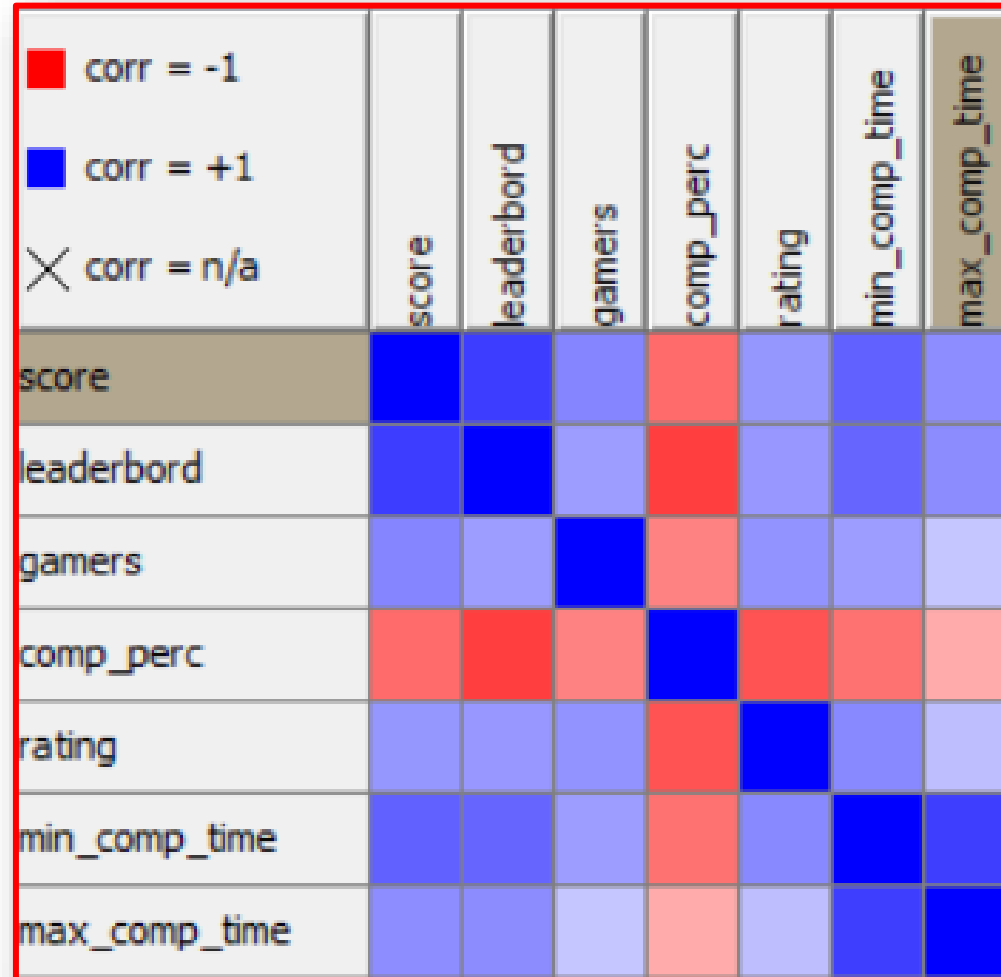
2 «Statistics» adlı node ile yüklediğimiz tablonun sütunlarındaki değerlerle ilgili betimleyici verilere ve histogramına ulaşabiliyoruz. Ayrıca sütunlarda eksik veri olmadığını da görebiliyoruz.



game score leaderbord gamers comp_perc rating url
No. missings: 0 No. missings: 0 No. missings: 0 No. missings: 0 No. missings: 0 No. missings: 0

min_comp_time max_comp_time
No. missings: 0 No. missings: 0

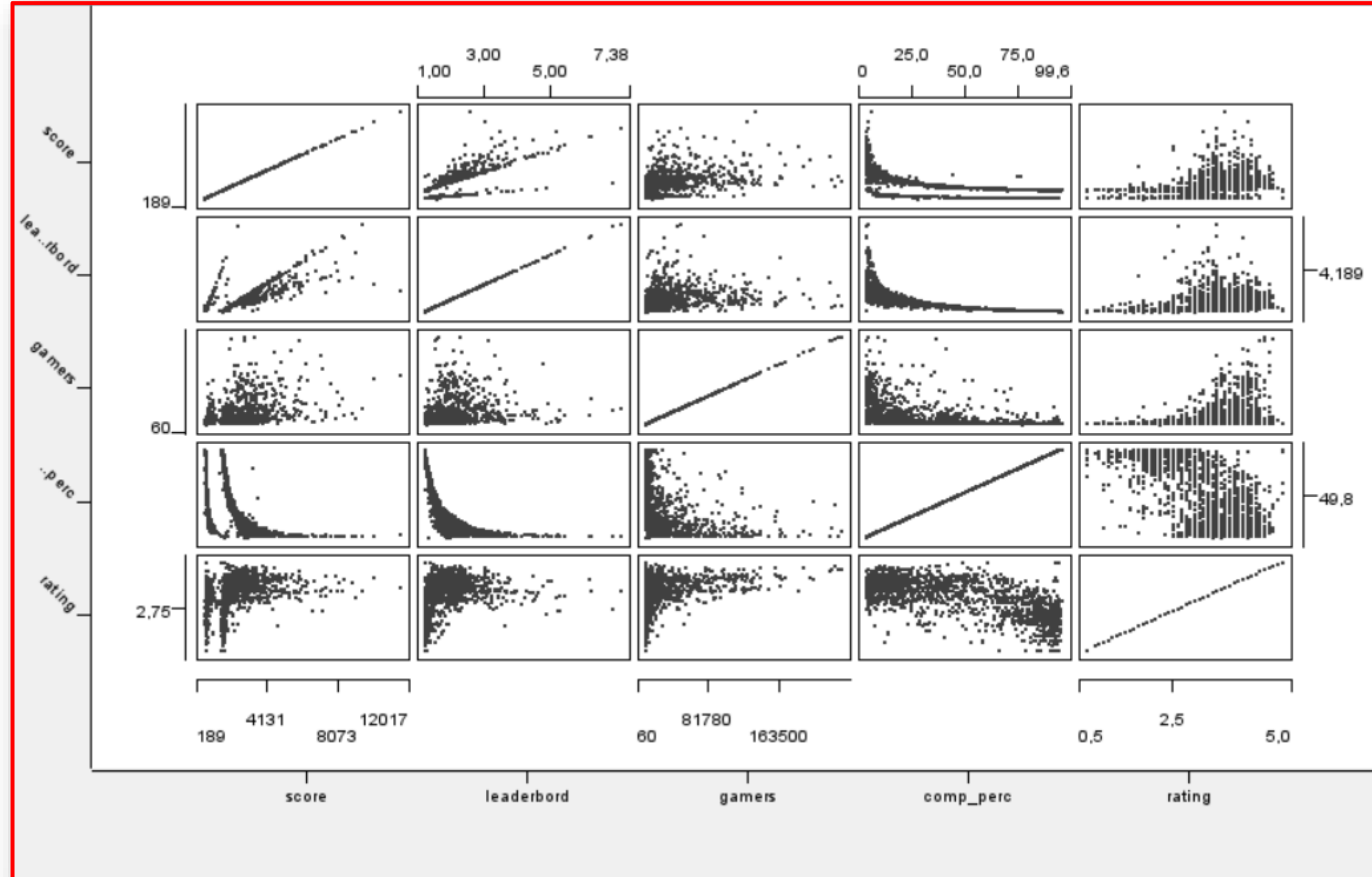
- 3 «Linear Correlation» adlı node ile yüklediğimiz tablonun sütunlarındaki değerler ışığında birbiri ile doğrusal ilişkide olanları veya ne ölçüde doğrusal ilişkiye sahip olduklarını görebiliyoruz. Soldaki görsel bunu renklerle kodlarken sayısal ilişkiyi sağdaki tabloda görebilmekteyiz.



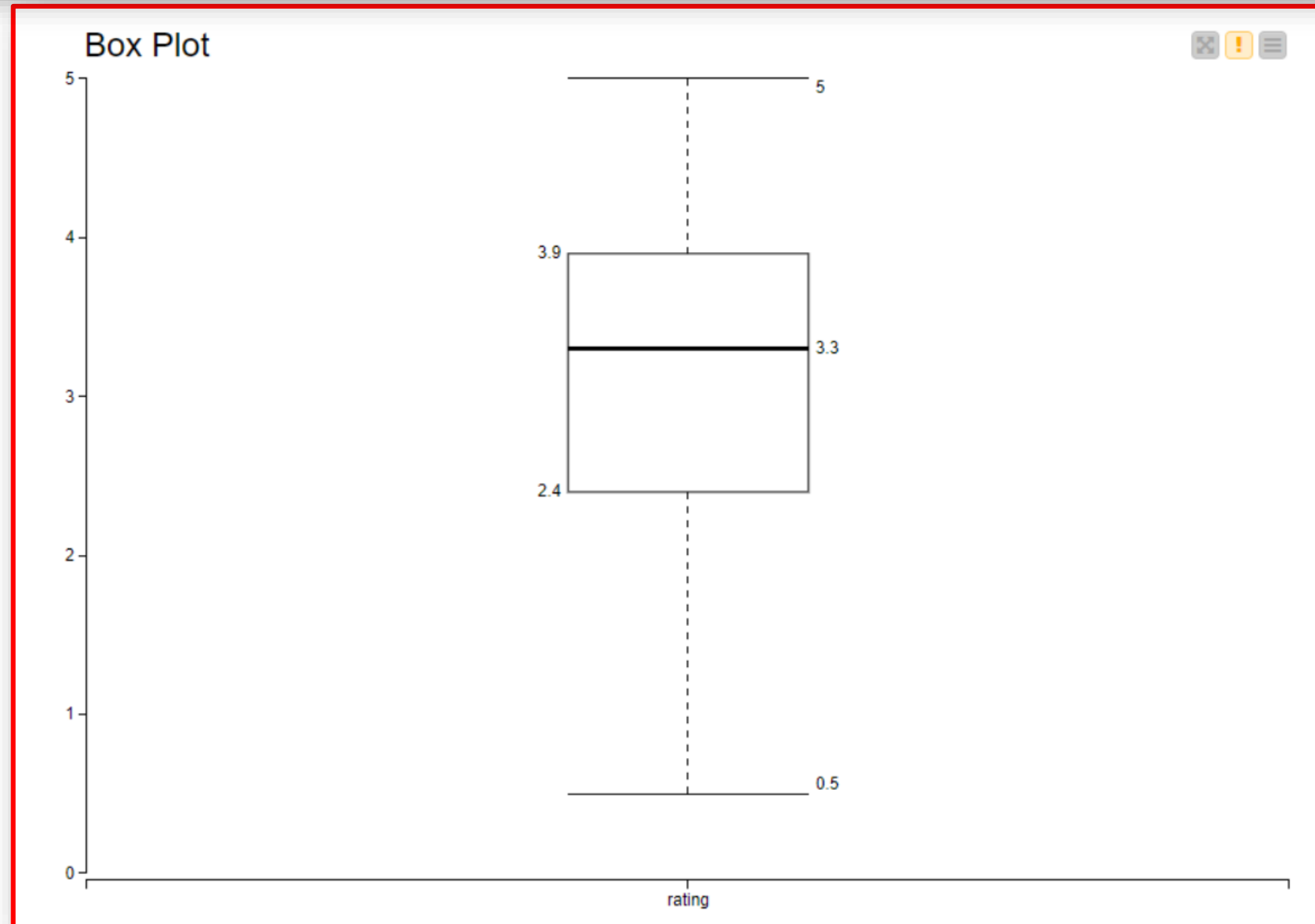
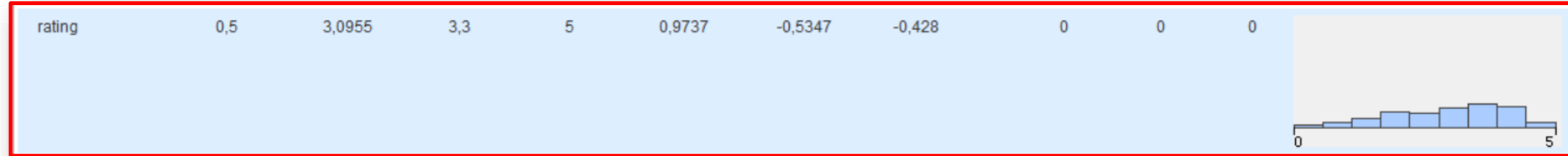
Tabloları incelediğimizde «score» ile «leaderboard» arasında doğrusal pozitif yönlü bir ilişki olduğunu yani birbirlerini doğru orantıyla etkileyen iki değişken olduklarını fark edebiliyoruz. Bununla birlikte «score» ile «comp-perc» arasında da doğrusal negatif yönlü bir ilişki olduğu görülmektedir. Özetle oyunda ulaşılan maksimum ganimet arttıkça oyunu tamamlayanların sayısı azlmaktadır.

File View							
Row ID	score	leaderbord	gamers	comp perc	rating	min comp t...	max comp ...
score	1	0.763	0.478	-0.58	0.411	0.621	0.448
leaderbord	0.763	1	0.385	-0.755	0.409	0.601	0.453
gamers	0.478	0.385	1	-0.49	0.423	0.385	0.227
comp_perc	-0.58	-0.755	-0.49	1	-0.68	-0.555	-0.328
rating	0.411	0.409	0.423	-0.68	1	0.464	0.256
min_comp_time	0.621	0.601	0.385	-0.555	0.464	1	0.755
max_comp_time	0.448	0.453	0.227	-0.328	0.256	0.755	1

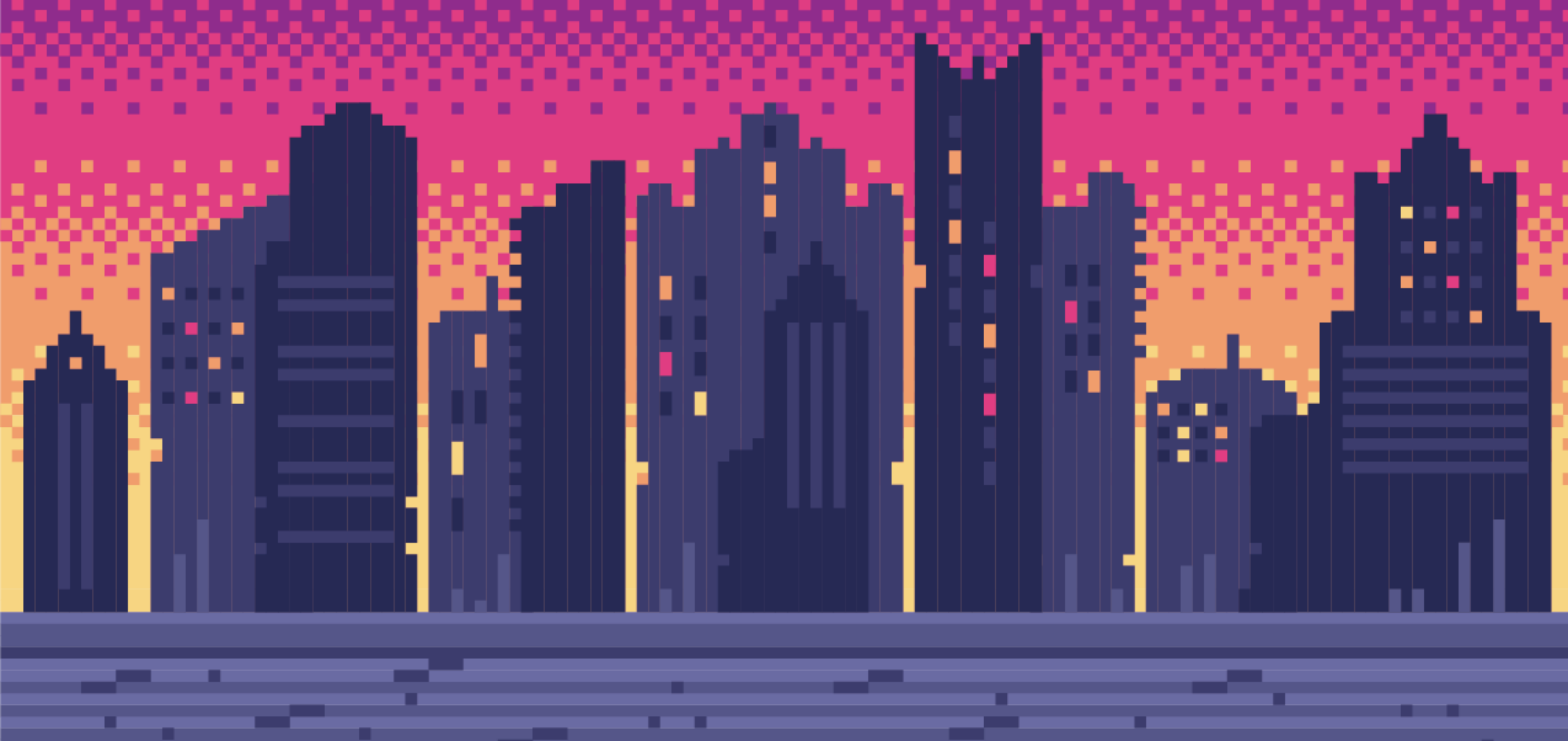
- 4 «Scatter Matrix» adlı komut ile yüklediğimiz tablonun sütunlarındaki değerler ışığında değişkenlerin biribiri ile ilişkisel dağılımlarını görebiliyoruz. Bir önceki yansıda «Linear Correlation» ile yaptığımız çıkarımları aşağıda daha somut olarak görebilmekteyiz.



- 5 «Statistics» komutuyla elde ettiğimiz temel betimleyici veriler «Box Plot» yardımı ile görselleştirildi. Bu görselleştirme normal dağılıma en yakın olan ve de problem durumunun odağında olacak «rating» değişkeni üzerinde yapıldı.



6. Verilerin anlamlandırılması, uygun problemin ne amaçla seçildiği üzerine:
«Oyunlar» ve «Oyuncu tipleri» ile ilgili literatürden bilgiler eklendi.



OYUNCU TİPLERİ

Oyuncu, herhangi bir oyunu oynayan kimsedir (TDK,2017). Bir diğer tanıma göre ise oyuncu bir oyundaki kuralları ve kısıtlamaları gönüllü olarak kabul eden kişidir. (Fullerton, 2008). Ancak bireylerin bu süreçte oynamayı tercih ettiği oyunlar ve oyun içerisindeki davranışları farklılaşabilmektedir. Dolayısıyla bireylerin beklentilerini karşılayan ve oynamak isteyecekleri oyunlar tasarlamaya ihtiyaç duyulmaktadır.





Bu durumda reklam, pazarlama gibi alanlarda tüketicilerin istek ve ihtiyaçlarıyla örtüşen ürün ve hizmet sunabilmek amacıyla yapılan sınıflandırma benzeri (Herbert, Charles, Moore ve Charles, 2014) bir yol izleyerek oyuncuların demografik-psikolojik özellikleri, oyundaki davranışları, beklentileri ve tercihleri gibi nitelikleri üzerinden sınıflandırılması, bireylerin motivasyonunun sağlanması ve memnuniyetlerinin artması noktasında faydalı olabilmektedir. (Bateman, Lowenhaupt ve Nacke, 2011; Bush vd., 2016; Crawford, 1984; Herbert vd., 2014; Juul, 2008; Neckle, Bateman ve Mandryk, 2011) Ayrıca oyuncuların farklı oyun tercihlerini belirlemek oyunun ticari açıdan da başarılı bir grafik çizmesini sağlamaktadır (Herbert, 2014; Nackle vd., 2011).



Dolayısıyla oyuncuları pek çok farklı açıdan inceleyen çalışmalar bulunmaktadır. Bu çalışmaların ilki Bartle'a (1996) ait olan ve oyuncuları;

- Kaşifler (explorers),
- Başaranlar (achievers),
- Sosyalleşenler (socializers) ve
- Katiller (killers)

Olarak dört kategoriye ayıran çok kullanıcıllı zindandır (multi-user dungeon, MUD).





Kaşifler

Oyunu keşfetmek, hataları bılmak, oyunun işleyişini anlamak gibi hedeflerle oyunda ilerler (Bartle, 1996; Samur, 2016).

Oyunda diğer oyuncularla iletişim kurup sosyalleşmeyi bilgi edinme yolu;puanı ise bir sonraki bölümde keşfedilmesi gerekenleri görmek için araç olarak görürler (Bartle, 1996)



Başaranlar

Oyunda verilen hedefe ulaşmak için her türlü riski alan oyuncu tipidir (Samur, 2016). Başaranlar, oyun içinde her türlü riski alırken oyun deneyimini geliştirmek için keşifler yaparlar, ganimetler toplar ve sosyalleşmeyi ise oyundaki ilerleyişlerine katkı sağlaması için bir yol olarak görürler (Bartle, 1996).



Sosyalleşenler

Sosyalleşenlerin temel amacıoyundaki kişilerle iletişime geçmektir (Samur, 2016). Bu oyuncu tipi tarafından oyun insanlarınıı buluşturan ortak payda ve puan ise statü kazanmanın bir yolu olarak kabul edilir. (Bartle, 1996).



Katiller

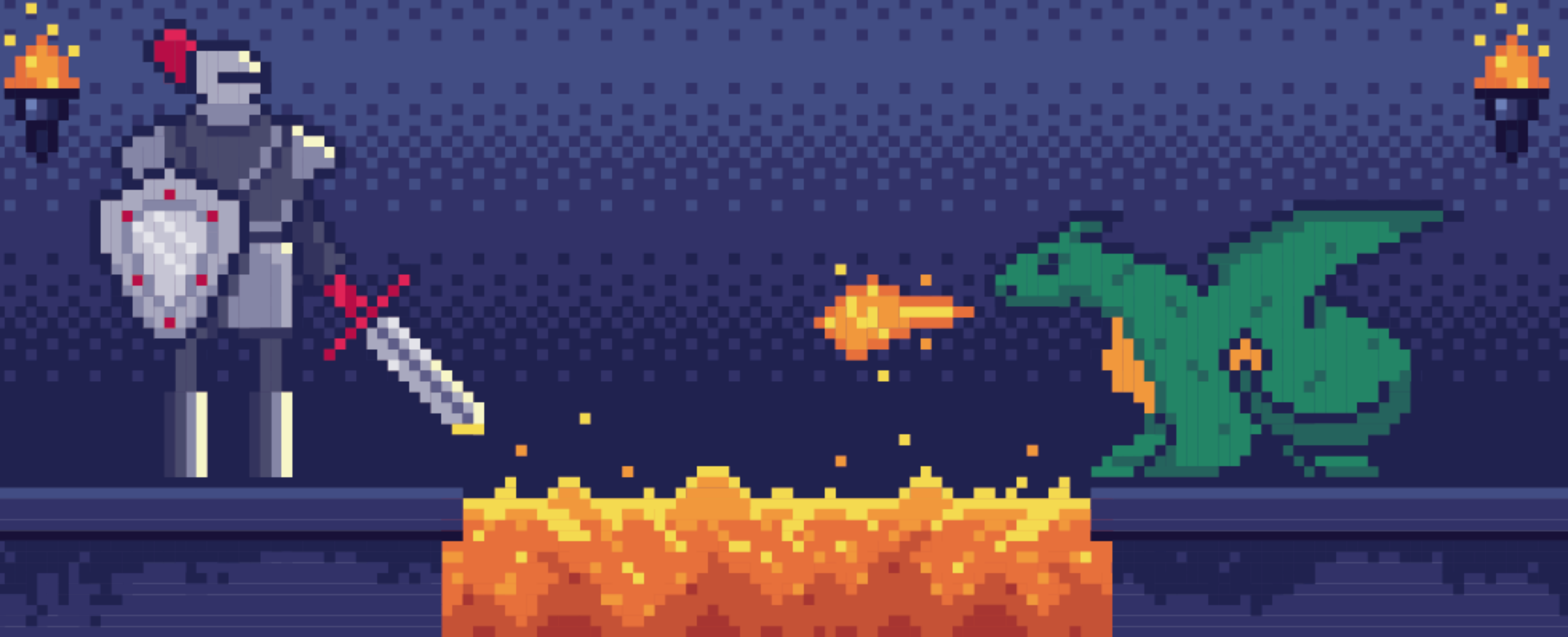
Katillerin oyundaki amacı oyunda kendi dışındaki tüm oyuncularını bir ödül beklentisi olmadan saf dışı bırakmak, elemektir (Bartle, 1996; Samur, 2016)

Sınırlılık



Farklı oyuncu tiplerinin oyundaki motivasyon kaynağı birbirinden farklı olmakla birlikte oyunlarda her oyuncu tipinin beklentisine uygun iletişim seçeneklerine, ödüle, merak uyandıracak nesnelere vb. yer verilmelidir (Bartle, 1996; William, Yee ve Caplan, 2008). Ancak Bartle'in yaklaşımında bireylerin farklı nedenlerle oyun oynadığı gerçeği ve benzer oyunların ya da bağlamların kişiler için farklı anlamlar taşıdığı göz ardı edilmektedir (Herbert vd., 2014; Yee, 2006). Ayrıca Bartle'in ortaya koyduğu modelde oyuncuların zamanla oyun içinde motivasyonunun değişebileceği de göz ardı edilmektedir (Herbert vd., 2014).

7. Veri setine uygun problem belirlendi.
8. Probleme uygun 3 model belirlendi.



Veriler incelendiğinde yapabileceklerimizi şöyle sıralayabiliriz.

1. Her bir değişkenin (sütunun) «**rating**»i nasıl etkilediği üzerine bir **Linear Regression** problemi kurulmuştur. Böylece uygun bir fonksiyon elde edilebilmiştir. En çok reytinge sahip oyunun en popüler olduğu varsayılmaktadır. Yapay Zeka öğrenmesi yöntemlerinden faydalanılmıştır.
(**Partitioning-> Linear Regression Learner->Linear Regression Predictor**)

➤ *Uzun vadede ise popülerliğe dayalı tavsiye sistemi oluşturulabilir.*

2. «**score**»lara göre oyunları «**Başaranlar**»/«**Achievers**» oyuncu tipine uygun olan veya olmayan olmak üzere bir sınıflandırma problemi kurulmuştur. **Logistic Regression ve Karar Ağacı** modelleri kullanılmıştır. Yapay Zeka öğrenme yöntemlerinden **K-fold Cross Validation** kullanılmıştır. Ayrıca iyileştirme için **Forward Feature Selection**'dan da faydalanılmıştır.

➤ *Uzun vadede ise oyuncu tiplerinin tamamının motivasyonları ve oyunların özellikleri göz önüne alınarak uygun oyunlar önerilebilir.*

Süreç Diyagramı

Problemin
Belirlenmesi

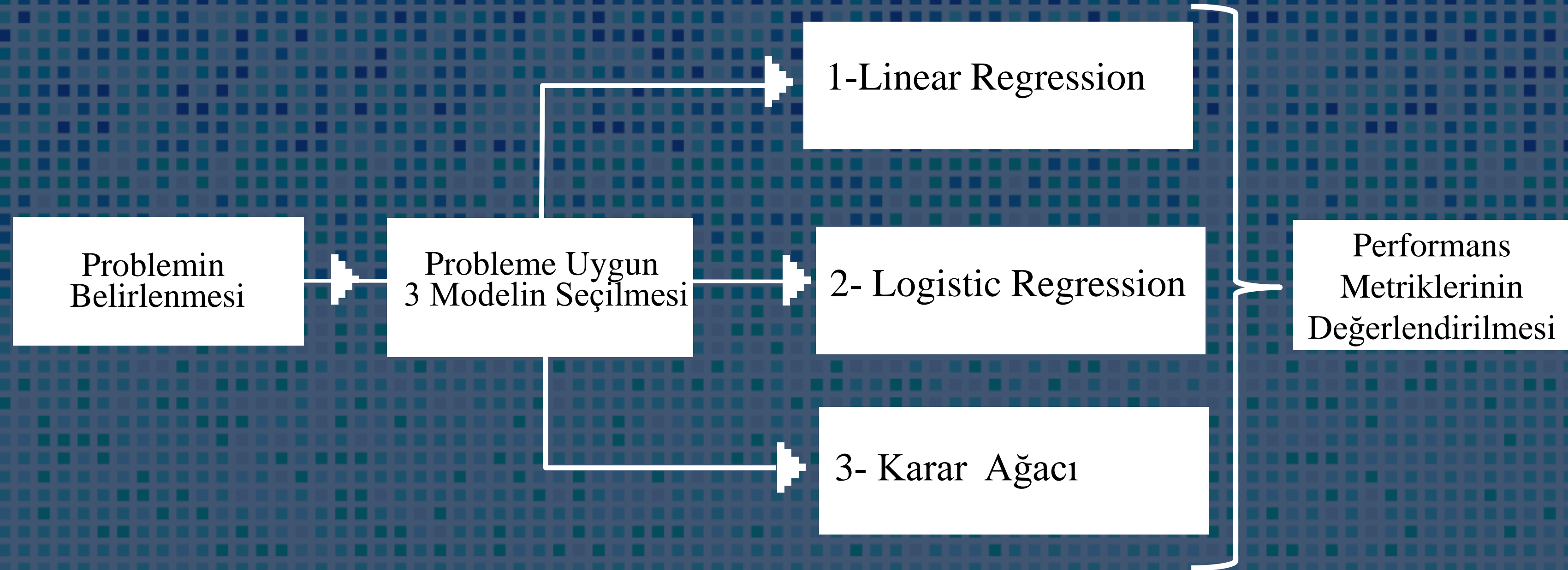
Probleme Uygun
3 Modelin Seçilmesi

1-Linear Regression

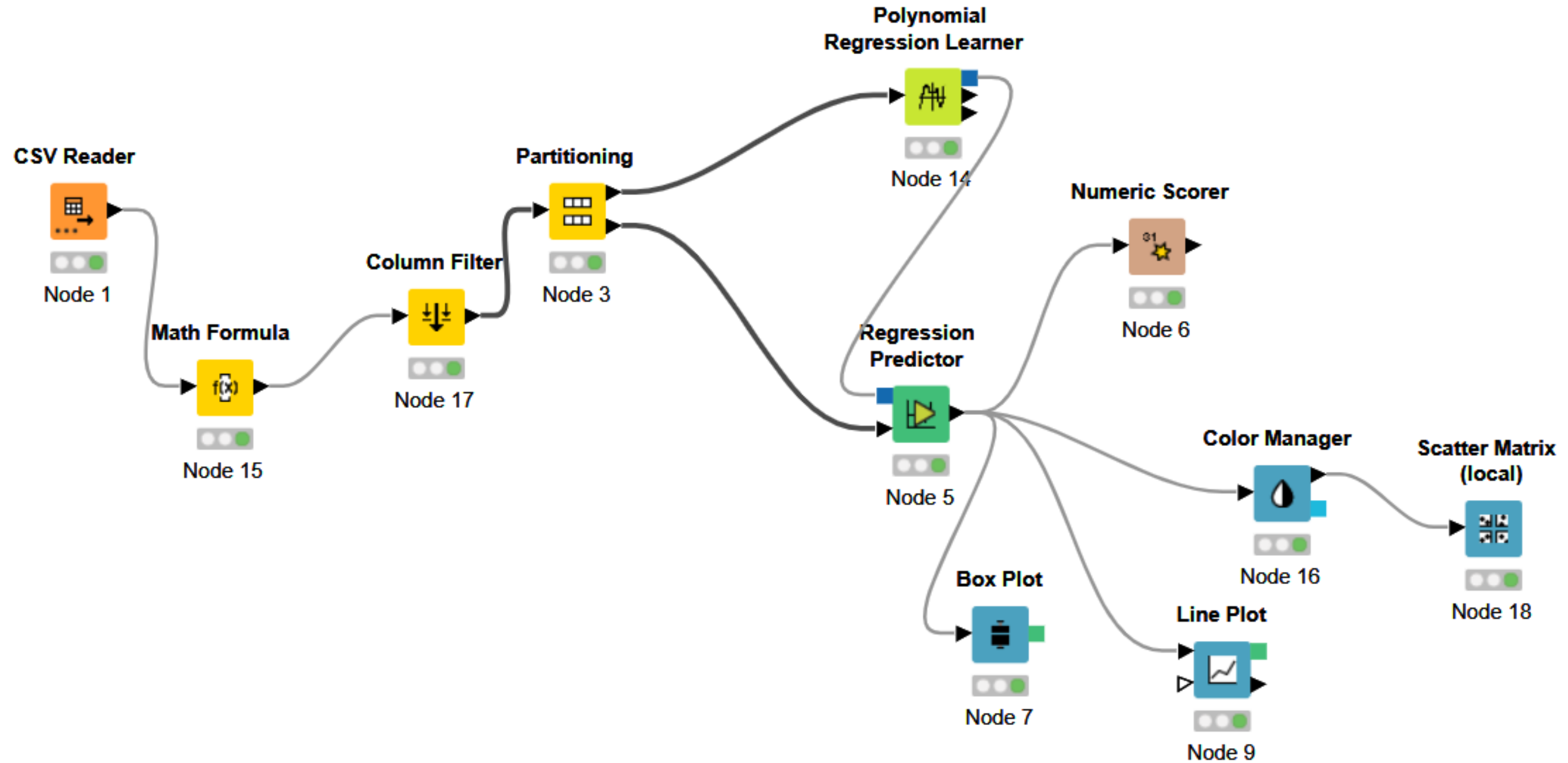
2- Logistic Regression

3- Karar Ağacı

Performans
Metriklerinin
Değerlendirilmesi



1-Linear Regression



1-Linear Regression

Adımları özetlemek gerekirse;

1. CSV dosyası yüklendi,
2. Ardından Linear regression için aşağıdaki değişkenler belirlenmiştir. Girdiler ve tahmin edilmesi beklenen «rating» değişkeni aşağıda gösterilmektedir.

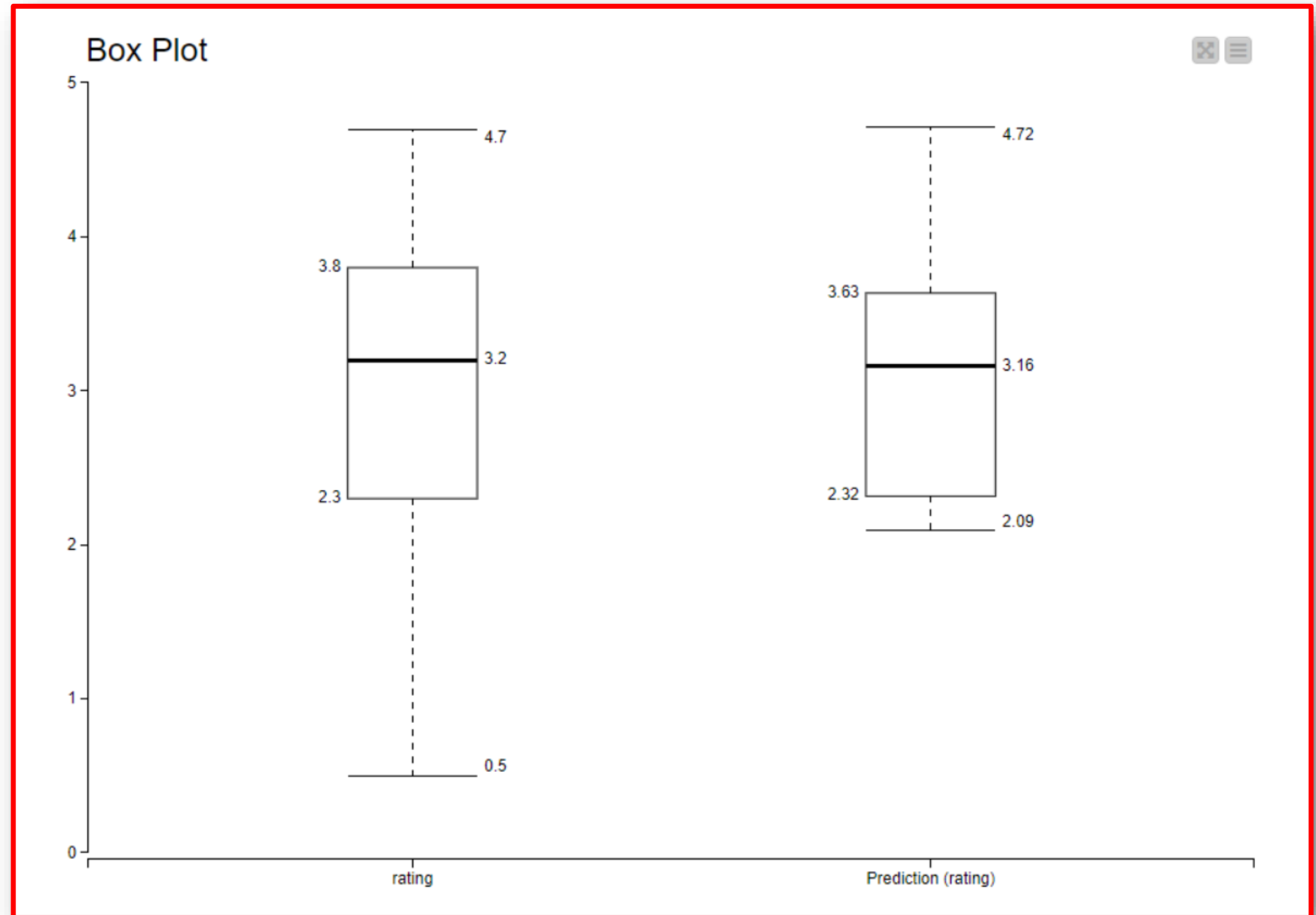
score (Number (integer))	x_1
leaderbord (Number (double))	x_2
gamers (Number (integer))	x_3
comp_perc (Number (double))	x_4
min_comp_time (Number (integer))	x_5
max_comp_time (Number (integer))	x_6

—————→ «rating»: $f(x)=y$

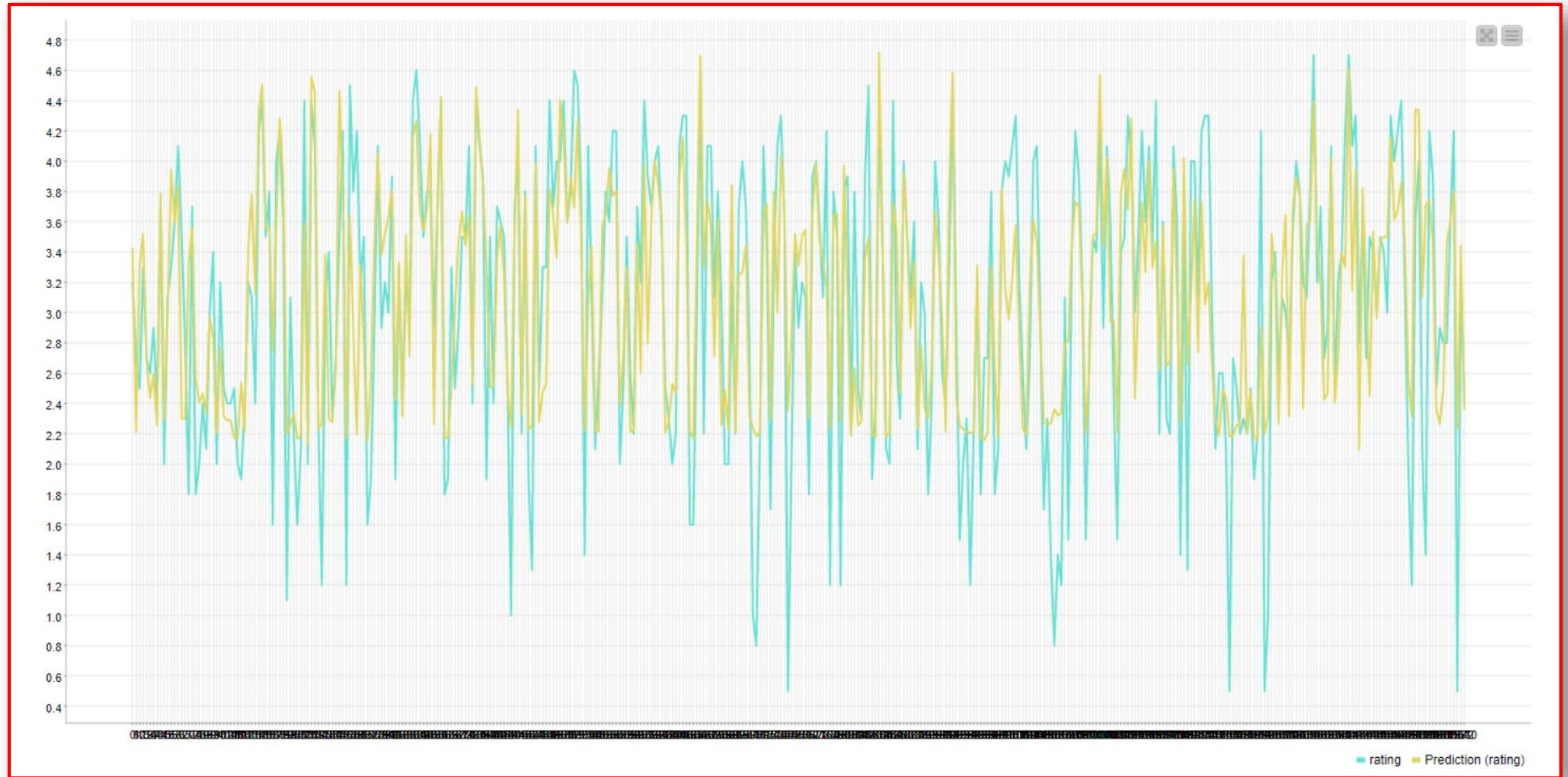
3. Partitioning ile %76 öğrenen %24 test olmak üzere bir parçalama uygulanmıştır.
4. Bu yapay zeka öğrenmesi sonucu performans metrikleri ve veri görselleştirmesi ise bir sonraki yansıda detaylandırılmıştır.

1-Linear Regression

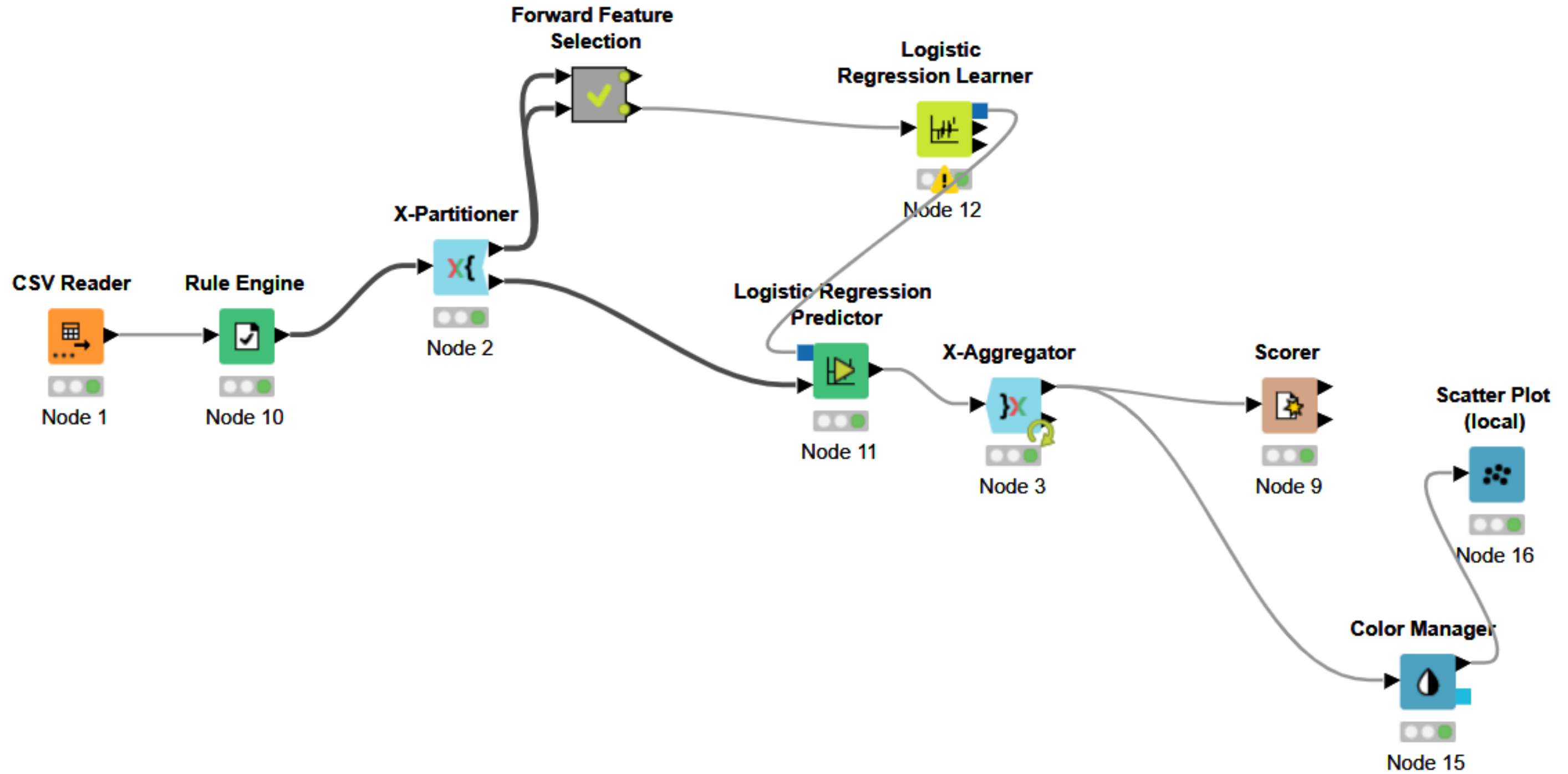
Stat...	—	□	×
File			
R ² :	0,624		
Mean absolute error:	0,438		
Mean squared error:	0,34		
Root mean squared error:	0,583		
Mean signed difference:	0,021		
Mean absolute percentage error:	0,213		
Adjusted R ² :	0,624		



1-Linear Regression



2-Logistic Regression



2-Logistic Regression

Adımları özetlemek gerekirse;

1. CSV dosyası yüklendi,
2. Ardından Rule Engine kullanılarak *KNIME Analytics* için sınıflama probleminde tahmin edilecek değişkenin kategorik olması gerektiğinden betimleyici adımlarda «score» değişkeninin histogramına uygun bir kategorizasyon yapılmıştır. Yeni bir sütun olarak «**Bartle Player**» veriye eklenmiştir.

```
$score$ >= 1224 => "High Achiever"  
$score$ < 1224 AND $score$ >= 300 => "Achiever"  
$score$ < 300 => "Low Achiever"
```

3. X-Partitioning ile Yapay Zeka öğrenmesi yöntemlerinden k-fold Cross Validation kullanılmış ve 5 gruba ayrılarak sınıflandırma modeli uygulanmıştır.
4. Süreçte performansı yükseltmek için «Forward Feature Selection» metanode uygulanmıştır. Böylece tüm yapay zeka öğrenmesi sürecinde en iyi performans elde edilmiştir.
5. Logistic regression performans skoru, ve veri görselleştirmesi ise bir sonraki yansıda detaylandırılmıştır.

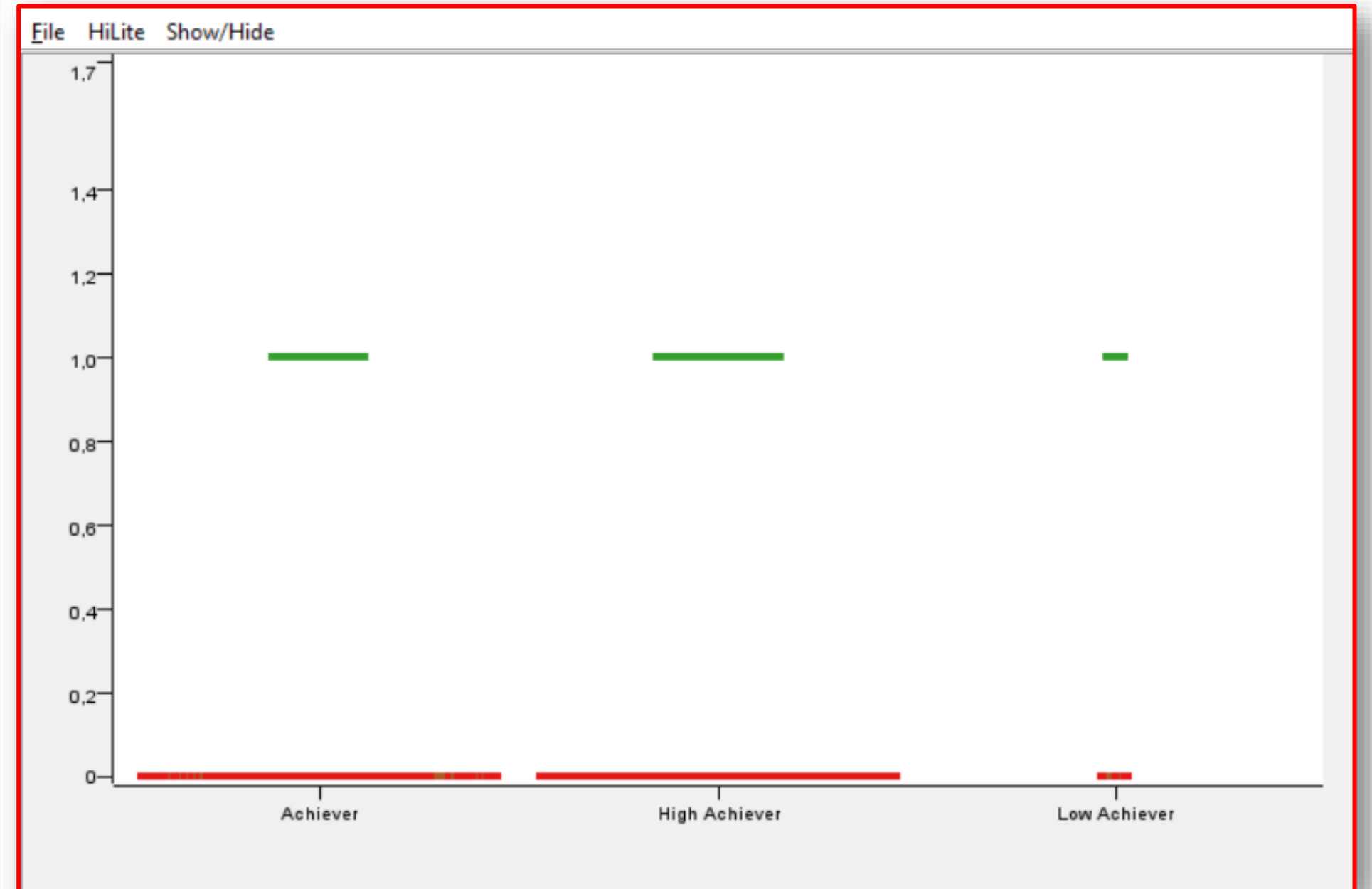
2-Logistic Regression

Confusion Matrix - 5:9 - Scorer

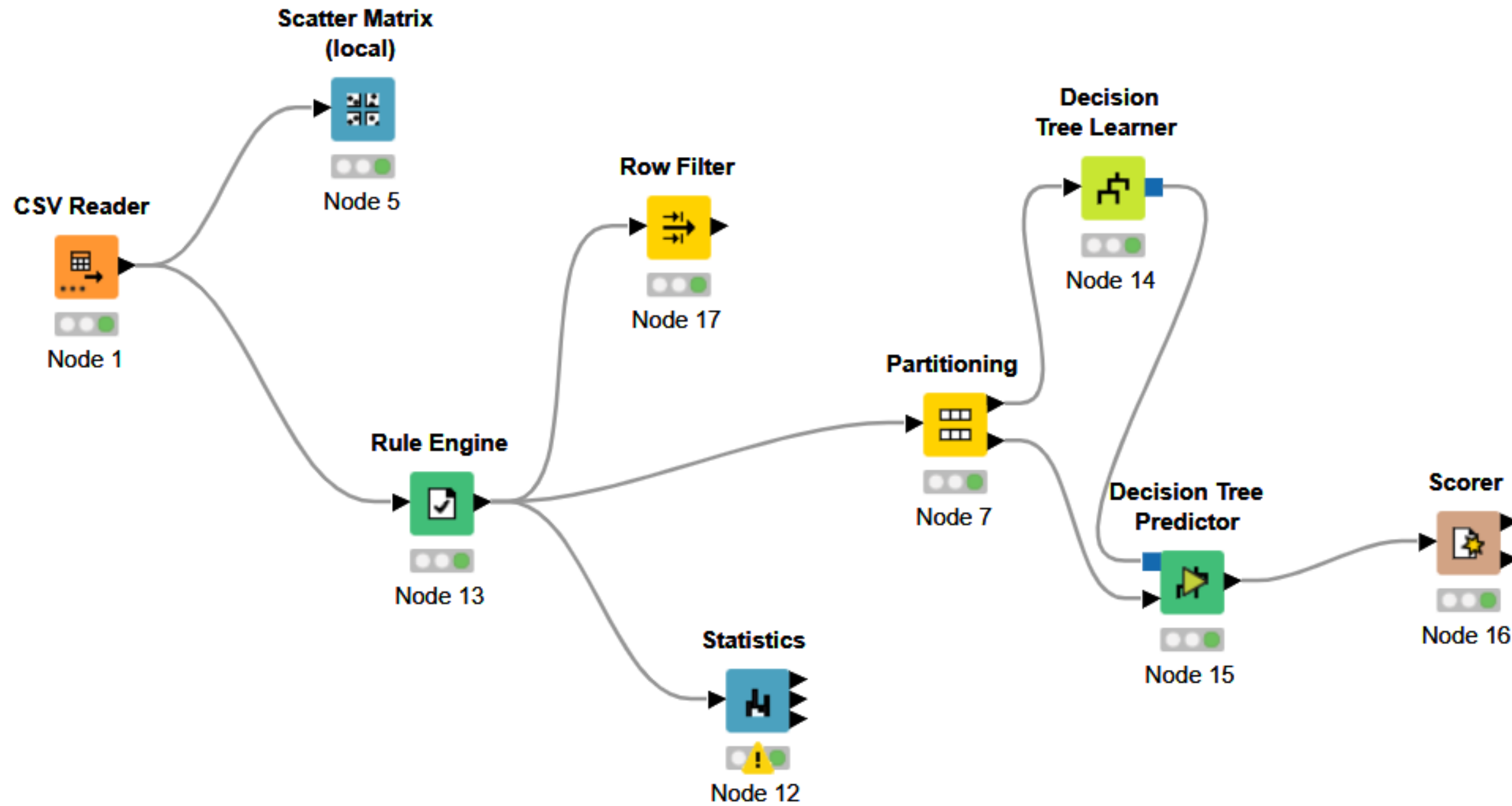
File Hilite

Bartle Play...	Achiever	High Achie...	Low Achiever
Achiever	52	232	16
High Achiever	69	1187	0
Low Achiever	11	12	5

Correct classified: 1.244
Wrong classified: 340
Accuracy: 78,535%
Error: 21,465%
Cohen's kappa (κ): 0,198%



3-Karar Ağacı



3-Karar Ağacı

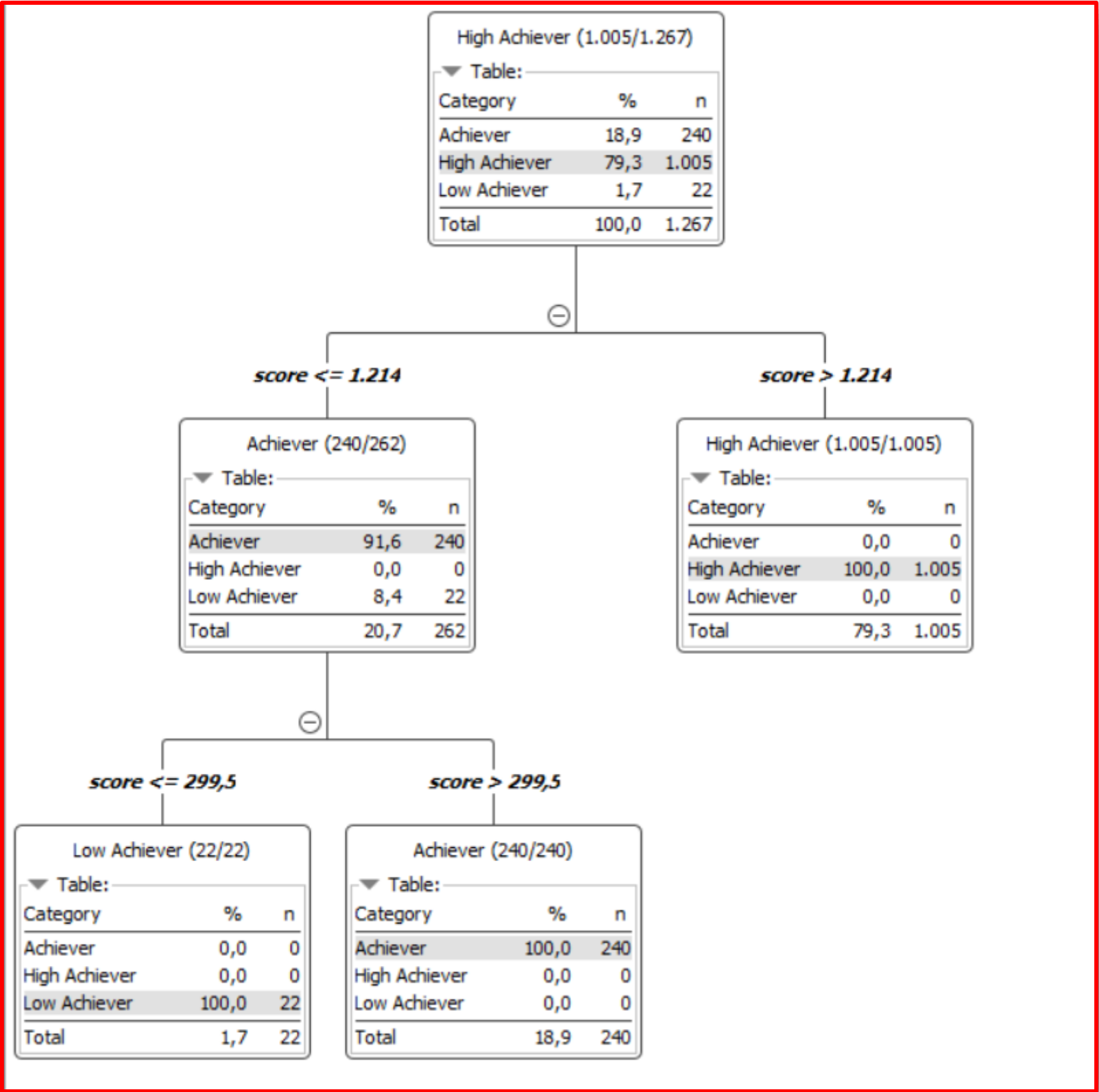
Adımları özetlemek gerekirse;

1. CSV dosyası yüklendi,
2. Ardından Rule Engine kullanılarak *KNIME Analytics* için sınıflama probleminde tahmin edilecek değişkenin kategorik olması gerektiğinden betimleyici adımlarda «score» değişkeninin histogramına uygun bir kategorizasyon yapılmıştır. Yeni bir sütun olarak «**Bartle Player**» veriye eklenmiştir. veriye eklenmiştir.

```
$score$ >= 1224 => "High Achiever"  
$score$ < 1224 AND $score$ >= 300 => "Achiever"  
$score$ < 300 => "Low Achiever"
```

3. Partitioning ile %80 öğrenen %20 test olmak üzere bir parçalama uygulanmıştır.
4. Bu yapay zeka öğrenmesi sonucu performans metrikleri ve veri görselleştirmesi ise bir sonraki yansıda detaylandırılmıştır.

3-Karar Ağacı



Sonuç ve Değerlendirme

Sonuçlar karşılaştırıldığında en iyi performansın sınıflama modellerinden karar ağacı olduğu gözükmemektedir.

Müşteri analitiği doğrultusunda sınıflama performansı geliştirilerek benzer özelliklere sahip oyunlar «Achievers»/ «Başaranlar» oyuncu tiplerine önerilebilir veya tahminin şaşırdığı ve olumlu olasılığın yüksek olduğu kesime bu oyunlar tekrar tavsiye olarak sunulabilir.

İlgili verilerle uzun vadeli yapılabilecek çalışmaları tekrar etmek gerekirse;

- Uzun vadede popülerliğe dayalı tavsiye sistemi oluşturulabilir.
- Uzun vadede oyuncu tiplerinin tamamının motivasyonları ve oyunların özellikleri göz önüne alınarak uygun oyunlar önerilebilir.

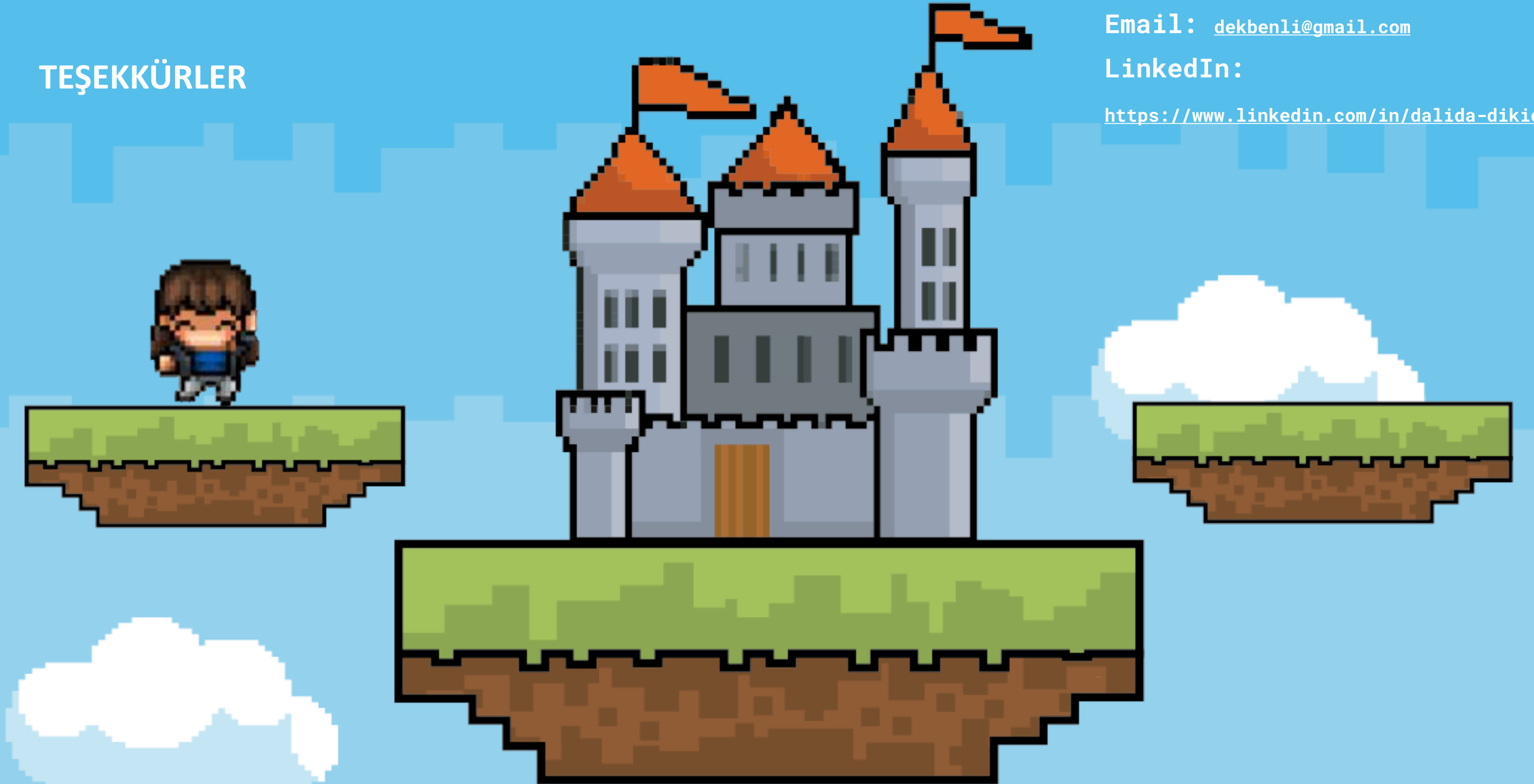
TEŞEKKÜRLER

Dalida Dikici

Email: dekbenli@gmail.com

LinkedIn:

<https://www.linkedin.com/in/dalida-dikici/>





GAME OVER