

Assistant Professor SELÇUK KIRAN

Data Mining und

Wissensentdeckung

In der Programmiersprache Python

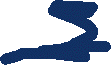
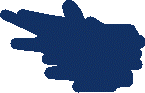
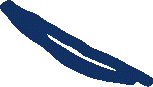
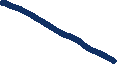
-CHI QUADRAT ANALYSE  
-LOGISTIC REGRESSION  
-RANDOM FOREST

**ULAŞ NAkİ TURAN-138417047**

**Selen Taşdelen-138418004**

2021

**Marmara Universität** **-**   **Wirtschaftsinformatik**



metin, duvar, iç mekan, kişi içeren bir resim

Açıklama otomatik olarak oluşturuldu**Ulaş Naki Turan (138417047)**

***Ich schwöre im Namen meiner Ehre und all der Werte, von denen ich geglaubt habe, dass ich bei dieser Prüfung gar nichts von irgendwo kopiert und eingefügt habe.***

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

\*\*\*Ich habe 50% zu dieser Aufgabe beigetragen.

duvar, kişi, iç mekan, kadın içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Selen Taşdelen (138418004)**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu***Ich schwöre im Namen meiner Ehre und all der Werte, von denen ich geglaubt habe, dass ich bei dieser Prüfung gar nichts von irgendwo kopiert und eingefügt habe.***

\*\*\*Ich habe 50% zu dieser Aufgabe beigetragen.

**1.UNSER ANALYSEZWECK**

**1.1 Chi-Quadrat Analyse**

Der Chi-Quadrat-Test wird verwendet, um zu analysieren, ob der Unterschied zwischen männlichen und weiblichen Studenten mit und ohne Berufserfahrung aus dem Datensatz statistisch signifikant ist. Chi-Quadrat-Test wird verwendet, da die Daten qualitativ mit Ja und Nein Mann und Frau angegeben werden .Beispielsweise kann die Hypothese, dass das Geschlecht keinen Einfluss auf die Berufserfahrung hat, je nach Häufigkeitsverteilung akzeptiert oder abgelehnt werden

**1.2 Logistische Regression**

Mit Hilfe der logistischen Regression wird es möglich sein, zu klassifizieren und zu bestimmen, ob 215 Studierende des Datensatzes je nach ihrer Qualifikation in einen beliebigen Arbeitsplatz eingeordnet werden können. An dieser Stelle werden die 14 Attribute des Datensatzes als unabhängige Variable charakterisiert, während der Status der Erwerbstätigkeit als abhängige Variable betrachtet wird und der Zusammenhang zwischen ihnen statistisch untersucht werden kann.

So liegt beispielsweise der prozentuale Erfolg einer Hochschulprüfung eines Studenten über dem Durchschnitt und es kann überprüft werden, ob sich dies auf den Stellenvermittlungsstatus auswirkt, indem 14 verschiedene Attribute wie Berufserfahrung usw. berücksichtigt werden. Da die abhängige Variable qualitativ bestimmt wird, wird der logistischen Regressionsanalyse der Vorzug gegeben.Mit der logistischen Regression wird versucht, die Situation zu modellieren, ob die Studierenden an einen Arbeitsplatz vermittelt werden oder nicht. Die logistische Regression umfasst binäre Daten für den Beschäftigungsstatus (1) und den Arbeitslosigkeitsstatus (0).

**1.3 Random Forest**

Unter Verwendung der 14 unabhängigen Attribute, die im Datensatz vorhanden sind, wird vom Random-Forest-Modell der Status, einen Job zu haben oder nicht zu haben, vorhergesagt.Das Ergebnis der Wirkung der unabhängigen Variablen auf die abhängige Variable wird als platziert oder in den Job platziert klassifiziert.Mit Hilfe der unabhängigen Variablen im Datensatz können Dutzende von Entscheidungsbäumen erstellt und diese Entscheidungsbäume trainiert werden.Anstatt von einem einzigen Entscheidungsbaum abhängig zu sein, werden viele Entscheidungsbäume erstellt und genauere Vorhersagen können gemacht werden.

Im Datensatz wurden sechzig Prozent der Schüler zum Trainieren des Algorithmus getrennt, während vierzig Prozent zum Testen aufgeteilt wurden.Nach dem Training kann eine Schätzung an den Testdaten durchgeführt werden. Eine hohe Genauigkeitsrate kann erreicht werden, indem die aus den Testdaten erhaltenen Werte und die vorhergesagten Werte verglichen werden.

**2.CAMPUS RECRUITMENT**

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

Der Datensatz enthält 215 Zeilen und 15 Spalten.Datensatz besteht aus Stellenvermittlungsdaten von Studenten auf dem Campus einer Universität in Indien.Der Erfolgsprozentsatz der Mittelschule, der Erfolgsprozentsatz der Oberstufe umfasst die Erfolgsquoten für Bachelor- und Masterstudium. Dazu gehört auch, ob die Schüler an der dem Zentrum angeschlossenen Schule studiert haben, ihr Geschlecht, ob sie über Berufserfahrung, ihre Fachgebiete und ihr Gehalt verfügen.

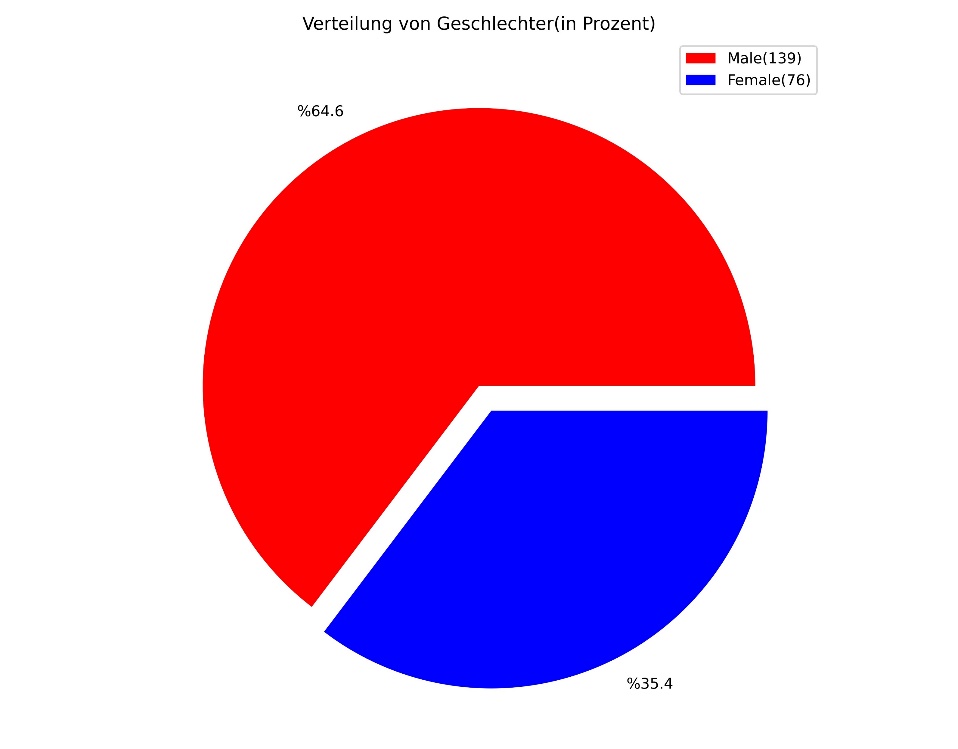
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

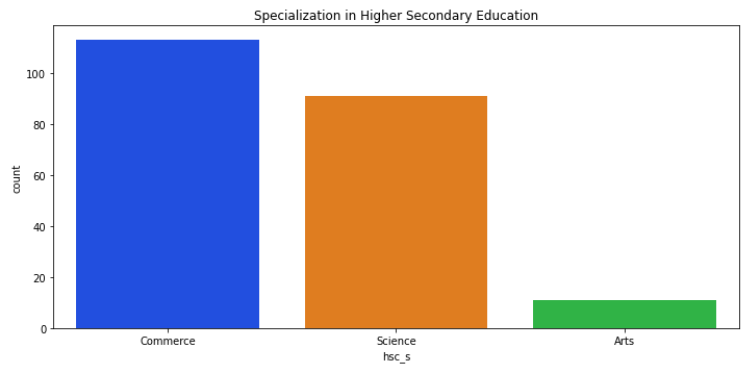
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

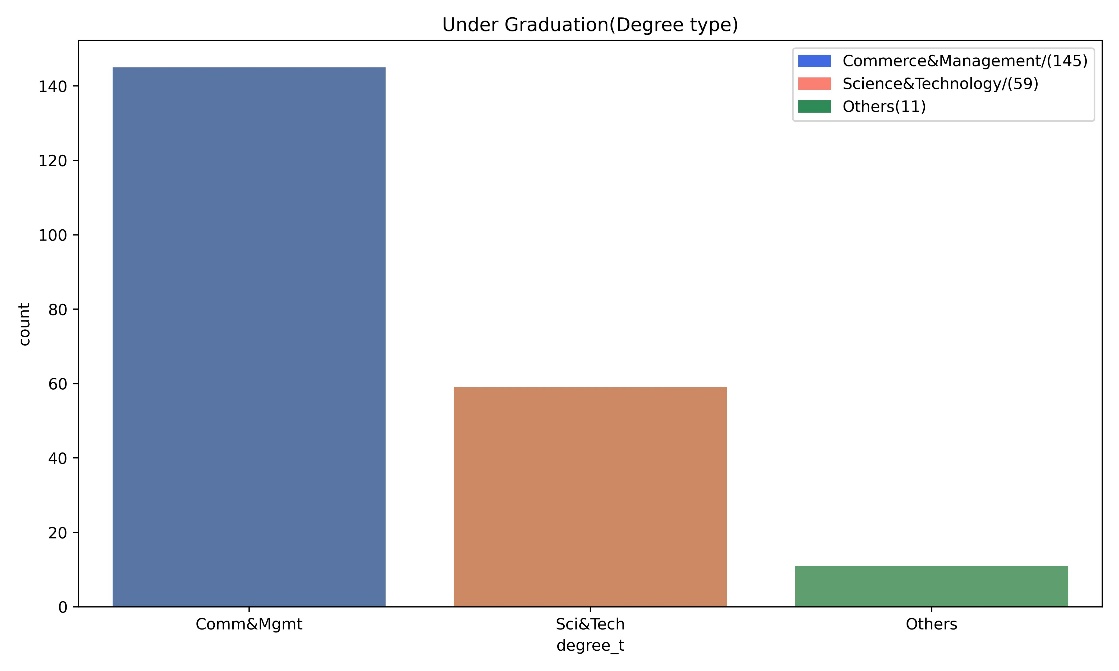
metin içeren bir resim

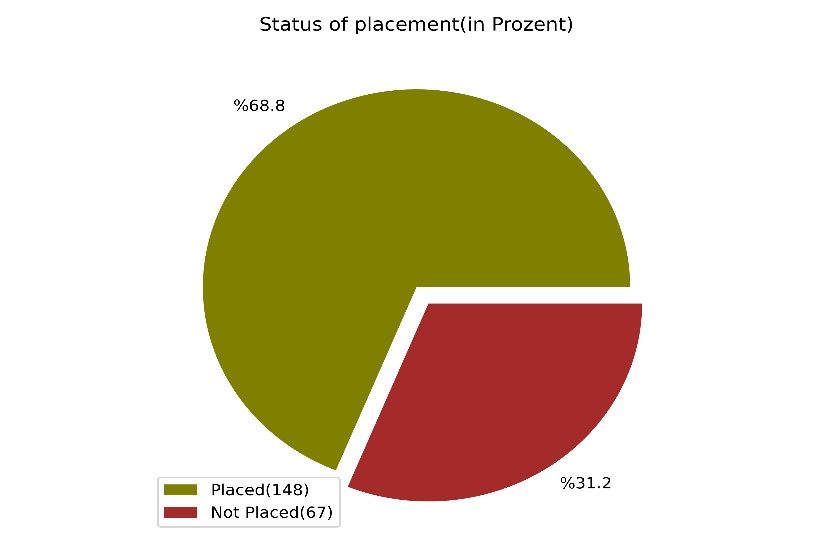
Açıklama otomatik olarak oluşturuldu**3.DATENVISUALISIERUNG**

**metin içeren bir resim

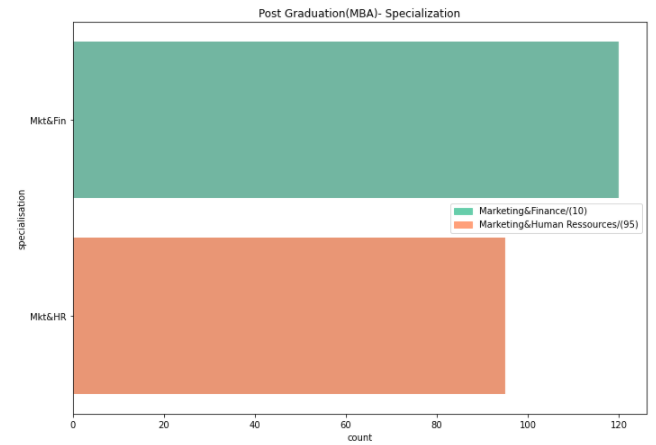
Açıklama otomatik olarak oluşturuldu**

**metin içeren bir resim

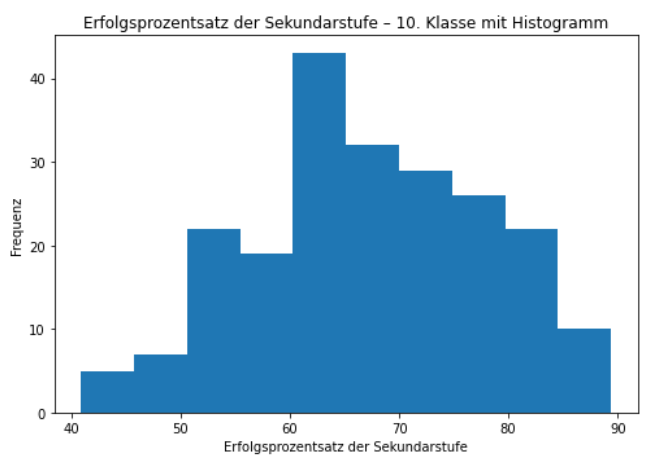
Açıklama otomatik olarak oluşturuldu**

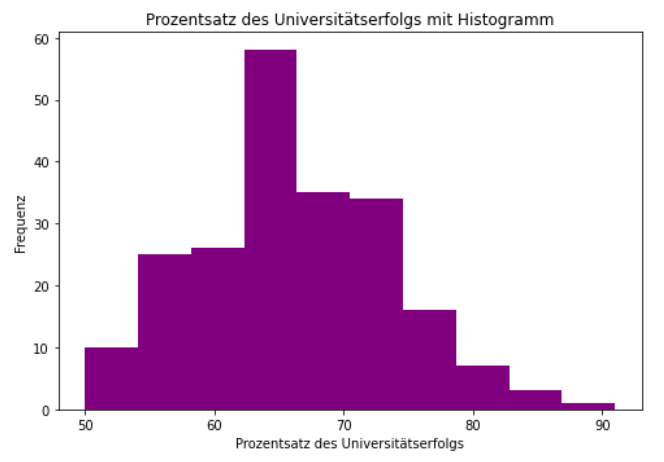
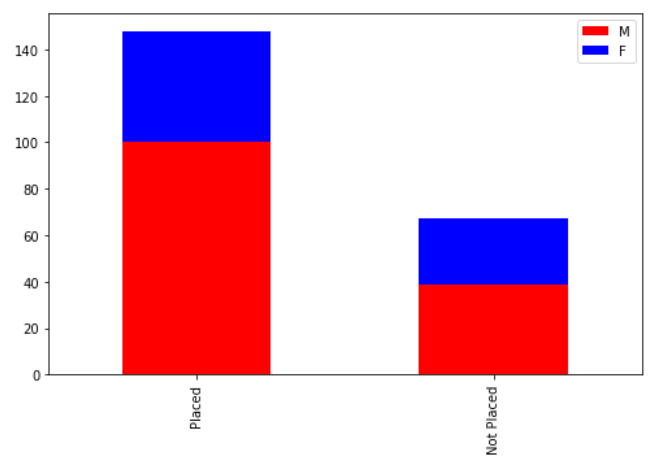
**metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

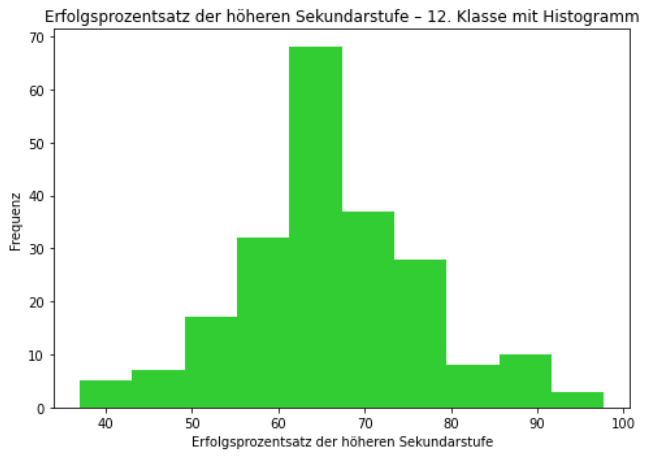
Açıklama otomatik olarak oluşturuldu**

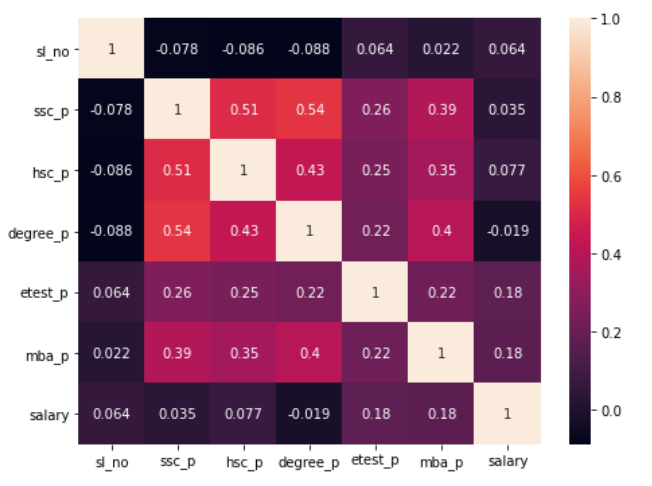
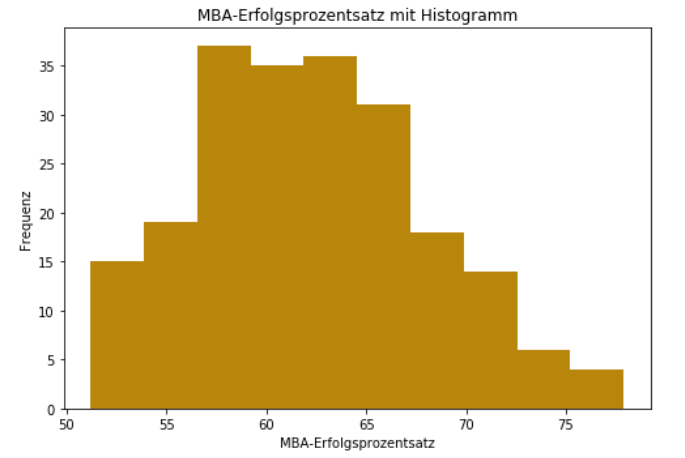
**metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

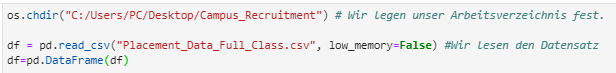
**metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu4.INSTALLATION DER BENÖTİGEN PAKETE**

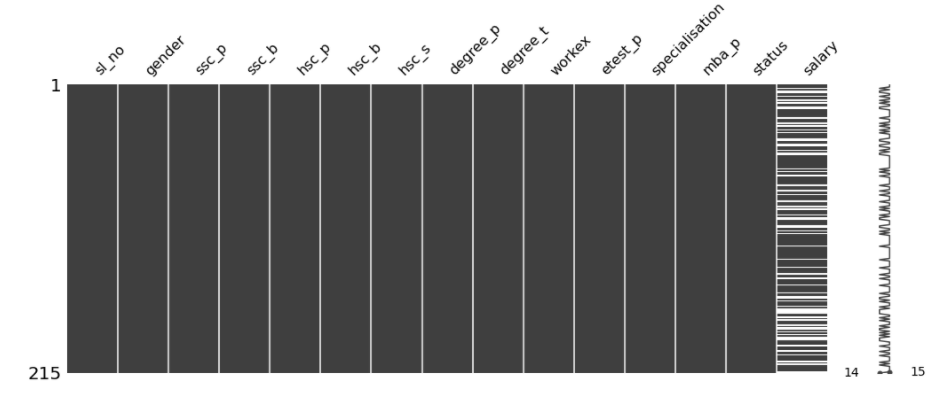
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldutablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**5.DATAREADING UND DATENSATZÜBERPRÜFUNG**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu**6.DATA PREPROCESSING**

****

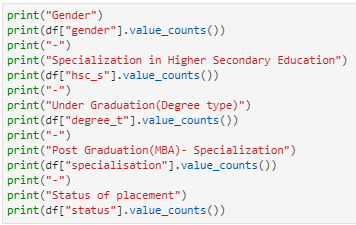
**7.EXPLORATIVE DATENANALYSE**

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldutablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**8. CHI QUADRAT ANALYSE**

H0: Es gibt keinen signifikanten Unterschied zwischen männlichen und weiblichen Studierenden mit Berufserfahrung.

metin içeren bir resim

Açıklama otomatik olarak oluşturulduH1: Es gibt einen signifikanten Unterschied zwischen männlichen und weiblichen Studenten mit Berufserfahrung.

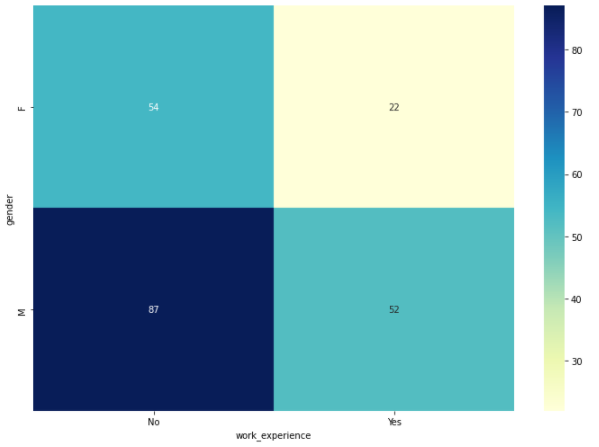
tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

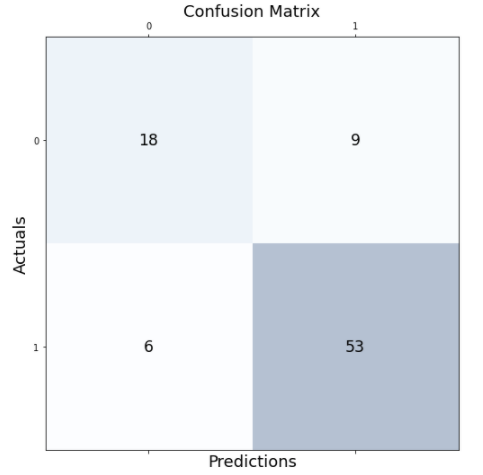
**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

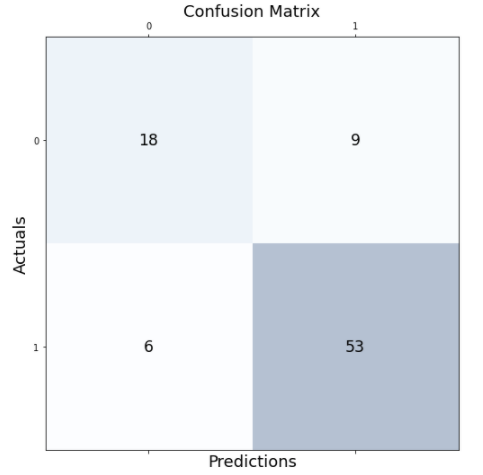
**metin içeren bir resim

Açıklama otomatik olarak oluşturuldumetin içeren bir resim

Açıklama otomatik olarak oluşturuldu9.LOGISTISCHE REGRESSION**

**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

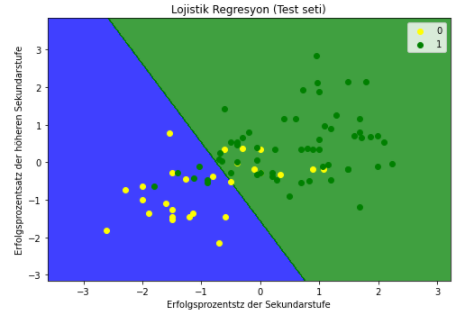


**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu



**metin içeren bir resim

Açıklama otomatik olarak oluşturuldu10. RANDOM FOREST**

**tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu**