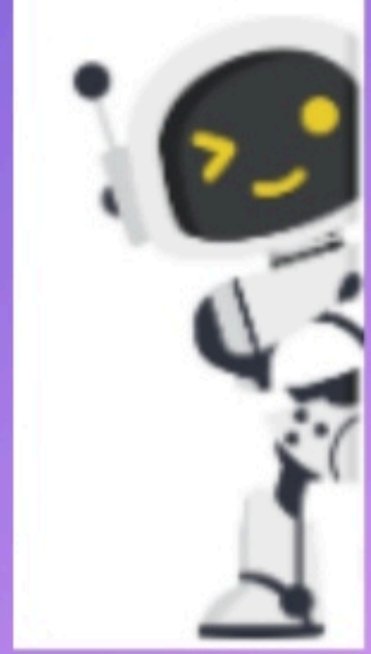


LOTUS AI

STAJ SÜRECİM

KÜBRA NUR BABACAN

LOTUS AI

EĞİTİM SÜRECİ

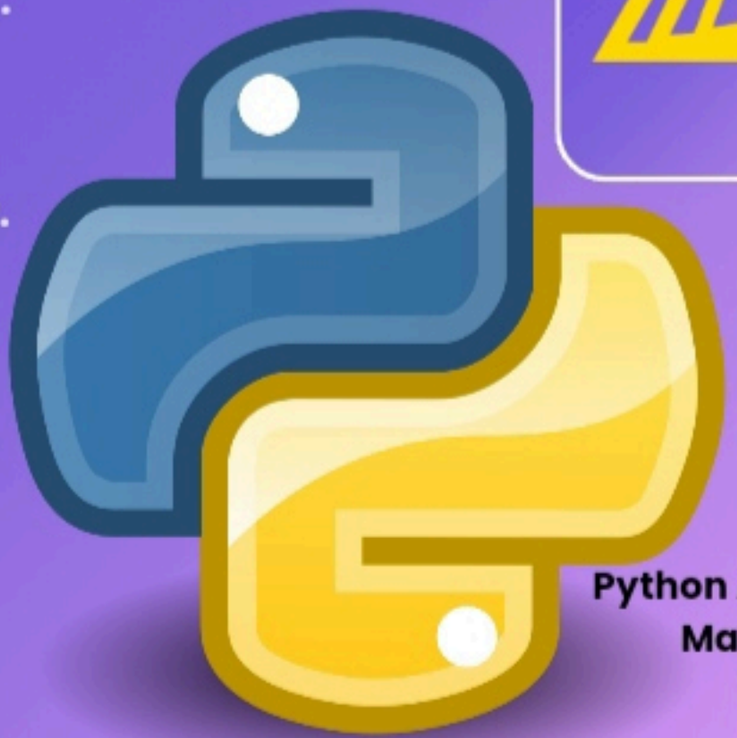
KNIME

- Kod yazmadan model oluşturma imkanı sağlıyor.
- Veri temizleme ve dönüştürme işlemleri pratik şekilde gerçekleşiyor.
- Raporlama ve analizde zaman tasarrufu sağlıyor.

PYTHON

- Veri bilimi ve makine öğrenmesinde geniş kütüphanelere sahiptir.
- Veri setlerini düzenleme ve manipüle etme imkanları vardır.
- Geniş görselleştirme kütüphanesi ile çeşitli görseller oluşturmak mümkündür.

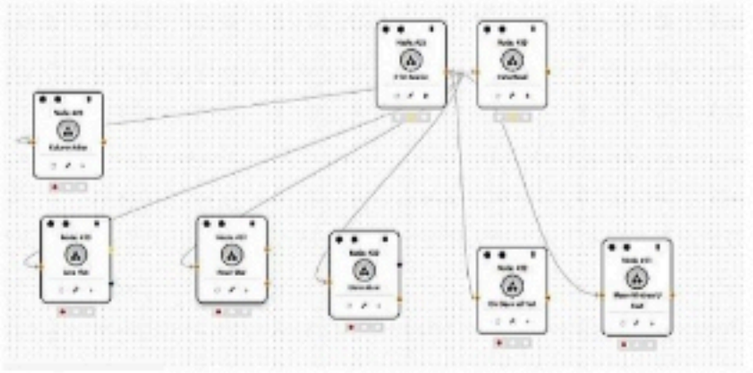
KNIME



Python A-Z™: Veri Bilimi ve Machine Learning (50 saatlk)

LOTUS AI

LOTUS AI DSML PLATFORMU

Ön Koşul:
• Kullanıcıların uygulamaya ya da web sitesine kayıtlı olabilmeleri için geçerli bir e-posta adresi ve şifre bilgilerine sahip olmaları gerekmektedir. • Kullanıcının / kullanıcıların üzerinde çalıştığı bir projeye sahip olması gerekmektedir.
Test Adımları:
• Kullanıcı e-posta adresini ve şifre bilgilerini doldurarak sisteme giriş yapmaktadır. • Kullanıcı çalıştığı projeyi açarak eklediği node'ları configurede hale getirir en son da reset all butonu ile çalışan tüm node'ları çalışmayı bekleyen node'ler haline (not configured) getirerek çalışmayı kaydetmesidir.
Beklenen: Çalışır durumda olan tüm node'ların resret all butonu ile çalışmayı durdurarak projenin hazır çalışmayı bekleme durumunda kalması beklenmektedir. **Mevcut:** Reset All butonumuz işlevini yerine getirmektedir.

Durum: **Başarılı**

General User Test (GUT)

- UI testler araştırılarak bilgi edinildi.
- Lotus AI Dsml platformu için genel kullanıcı testleri gerçekleştirilmiştir.

LOTUS AI

LOTUS AI

100 ETL (Extract, Transform, Load)

```sql
92) En çok kazanç sağlayan film ve kategorisi;
SELECT film.title, category.name AS genre, SUM(payment.amount) AS total_revenue
FROM film
JOIN film_category ON film.film_id = film_category.film_id
JOIN category ON film_category.category_id = category.category_id
JOIN inventory ON film.film_id = inventory.film_id
JOIN rental ON inventory.inventory_id = rental.inventory_id
JOIN payment ON rental.rental_id = payment.rental_id
GROUP BY film.title, category.name
ORDER BY total_revenue DESC
LIMIT 5
```

- 100 tane ETL soruları hem SQL ile hem de PYTHON / Pandas ile analiz edilerek çıkarılan sonuçlar karşılaştırılmıştır.

PANDAS - SQL

- **SQLite ve Python / JupyterLab kullanılmıştır.**

```python
import pandas as pd
import sqlite3
# sqlite ile veritabanına bağlanıyoruz.
conn = sqlite3.connect(file_path)
# dataframe listesi oluşturma
df = {}
for table_name in ['film', 'film_category', 'category', 'inventory', 'rental', 'payment']:
    query = f"SELECT * FROM {table_name}"
    df[table_name] = pd.DataFrame(conn.execute(query).fetchall(), columns=[column[0]
                                                                           for column in conn.execute(query).description])
max_kazanclı_film_kategorisi = df['film'] \
    .merge(df['film_category'], on='film_id', suffixes=('', '_fc')) \
    .merge(df['category'], on='category_id', suffixes=('', '_cat')) \
    .merge(df['inventory'], on='film_id') \
    .merge(df['rental'], on='inventory_id') \
    .merge(df['payment'], on='rental_id', suffixes=('', '_pay')) \
    .groupby(['title', 'name']) \
    .agg(total_revenue=('amount', 'sum')) \
    .reset_index() \
    .sort_values(by='total_revenue', ascending=False) \
    .head(5)
print("En çok kazanç sağlayan film ve kategorisi: ")
print(max_kazanclı_film_kategorisi)
```

```
En çok kazanç sağlayan film ve kategorisi:
                title         name  total_revenue
841   TELEGRAPH VOYAGE        Music         231.73
950         WIFE TURN  Documentary         223.69
```

LOTUS AI

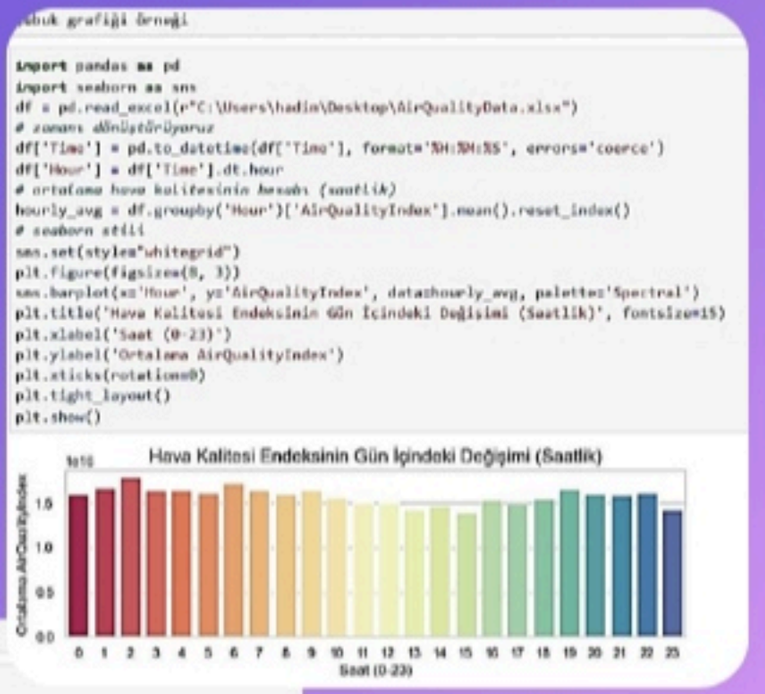
LOTUS AI

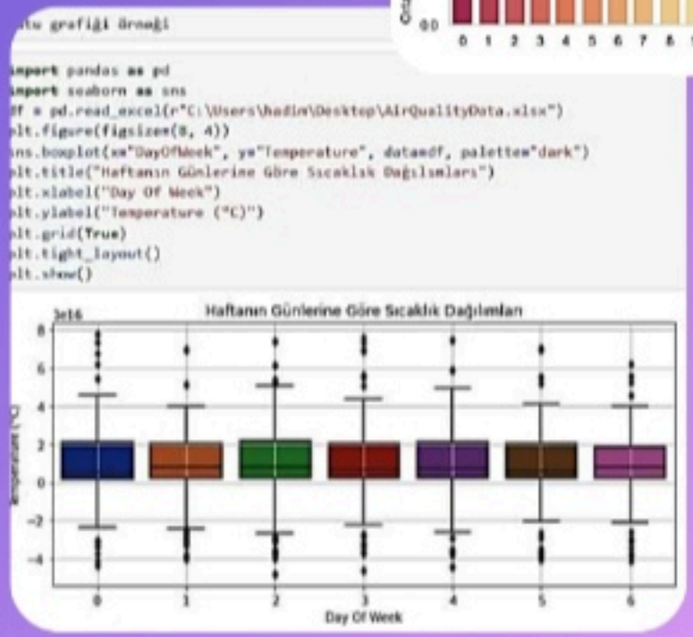
VERİ GÖRSELLEŞTİRME

- Grafik türleri , avantaj / dezavantajı , kullanım amacı , veri görselleştirme kütüphaneleri hakkkında araştırma yapılmıştır.

- Araştırma sonrasında görselleştirme işlemleri hem KNIME ile hem de PYTHON ile görselleştirilmiştir.

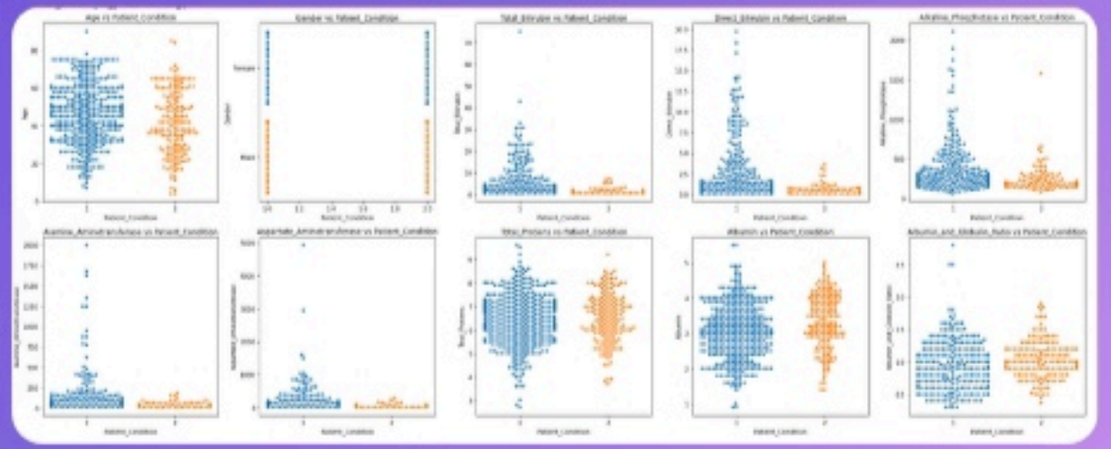
- Hava kalitesi verileri ile analiz edilmiştir.





LOTUS AI

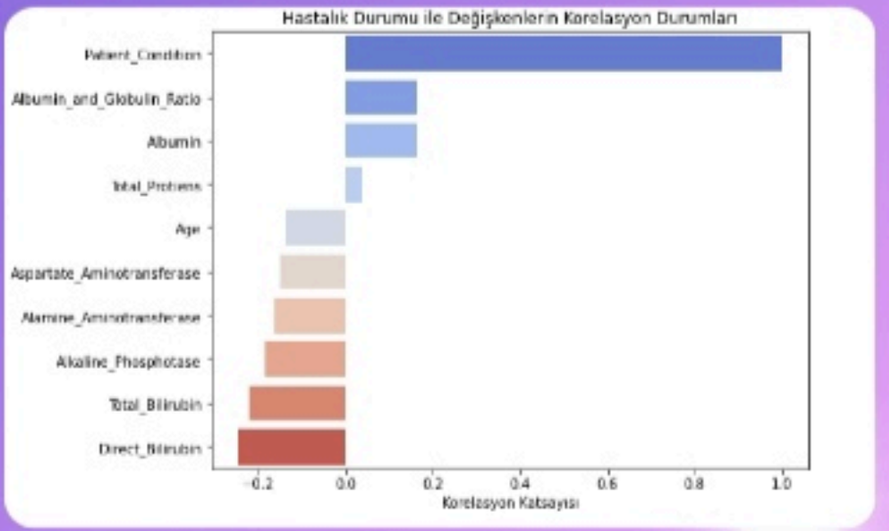
5 MAKİNE ÖĞRENMESİ PROJESİ



- Sınıflandırma Analizi
- Anamoli Tespiti
- Regresyon Analizi
- İlişki Kuralı Analizi
- Kümeleme Analizi

KARACİĞER HASTALIĞI OLAN VE OLMAYAN BİREYLER İÇİN SINIFLANDIRMA ANALİZİ

- Keşifsel veri analizi (EDA) işlemleri gerçekleştirilmiştir.
- Veriler görselleştirilmiştir.
- Makine öğrenmesi analizi gerçekleştirilip sonuçlar yorumlanmıştır.



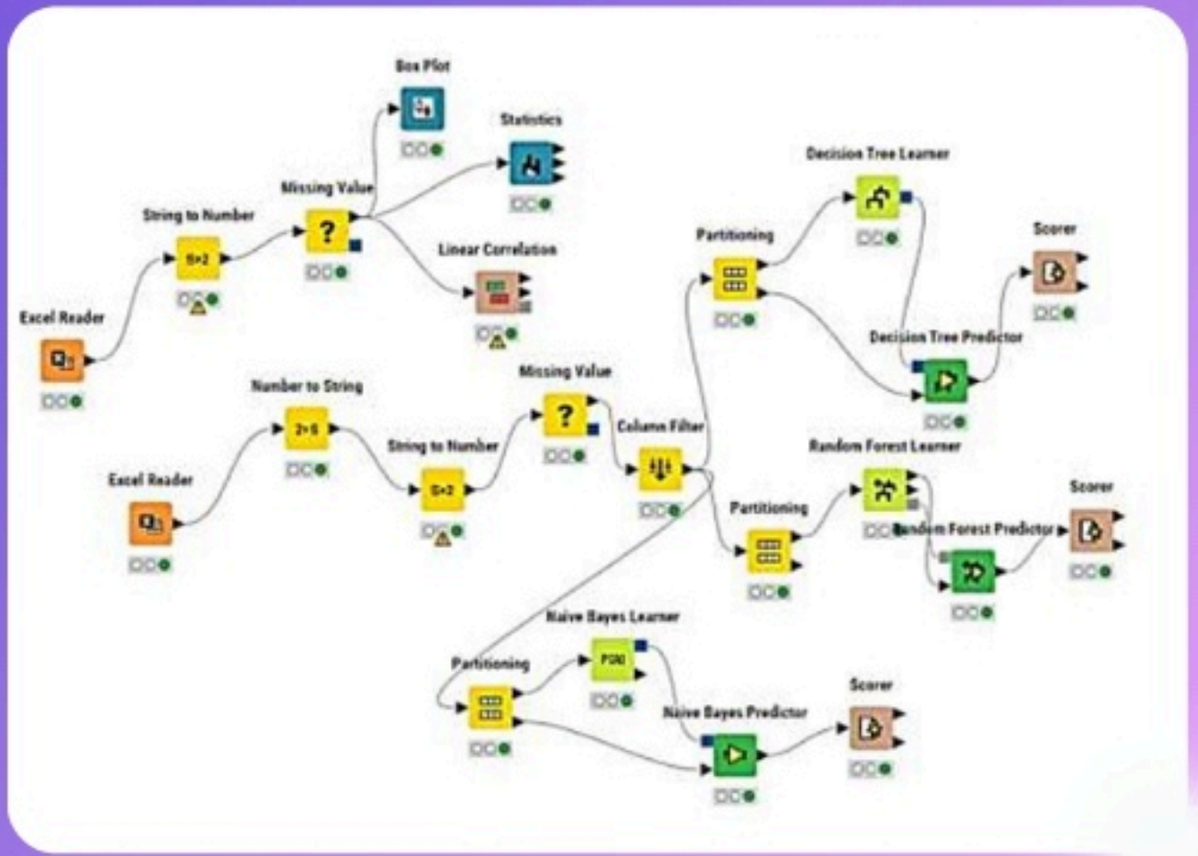
LOTUS AI

LOTUS AI

5 MAKİNE ÖRĞENMESİ PROJESİ

KNIME ANALİZİ

- Karaciğer hastalığı olan ve olmayan bireyler için PYTHON'dan sonra KNIME analizi yapılmıştır.
- Gerekli node'lar eklenip düzenlemeler yapıldıktan sonra analiz sonuçları karşılaştırılmıştır.



LOTUS AI

LOTUS AI

5 MAKİNE ÖRĞENMESİ PROJESİ

YETİŞKİN NÜFUS SAYIMI GELİRİ ANALİZİ

- **Keşifsel veri analizi (EDA) işlemler gerçekleştirilmiştir.**

- **ANAMOLİ TESPİTİ**

```
Anamoli Tespiti

from sklearn.ensemble import IsolationForest
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# veriyi hazırlama
columns_to_drop = []
if 'fnlwgt' in data.columns:
    columns_to_drop.append('fnlwgt')
if 'income' in data.columns:
    columns_to_drop.append('income')

X = data.drop(columns=columns_to_drop)
y = data['income']

# anomali tespiti
iso_forest = IsolationForest(contamination=0.05, random_state=42)
outliers = iso_forest.fit_predict(X_train)

# anomali olanları çıkarma temizleme
X_train_clean = X_train[outliers == 1]
y_train_clean = y_train[outliers == 1]

# model denemesi
model = LinearRegression()
model.fit(X_train_clean, y_train_clean)

y_pred = model.predict(X_test)

# değerlendirme
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print(f"Test Mean Squared Error (MSE): {mse:.4f}")
print(f"Test R-squared (R²): {r2:.4f}")
```

- **REGRESYON ANALİZİ**

```
Model: Linear Regression
Best Parameters: {'fit_intercept': True}
Test MSE: 0.1408
Test R²: 0.2192

Model: Ridge
Best Parameters: {'alpha': 10}
Test MSE: 0.1408
Test R²: 0.2192

Model: Lasso
Best Parameters: {'alpha': 0.1}
Test MSE: 0.1485
Test R²: 0.1765

Model: ElasticNet
Best Parameters: {'alpha': 0.1, 'l1_ratio': 0.1}
Test MSE: 0.1418
Test R²: 0.2134

Model: Decision Tree
Best Parameters: {'max_depth': 10, 'min_samples_split': 10}
Test MSE: 0.1224
Test R²: 0.3209

Model: Random Forest
Best Parameters: {'max_depth': 10, 'n_estimators': 200}
Test MSE: 0.1125
Test R²: 0.3760

Model: Gradient Boosting
Best Parameters: {'learning_rate': 0.1, 'n_estimators': 200}
Test MSE: 0.1119
```

- **ANAMOLİ TESPİTİ**
  Değişken çıkarılması ve anamoli olan değerlerin tespiti sağlanmıştır.
- **REGRESYON ANALİZİ**
  Linear regresyon , Ridge regresyon , Lasso regresyon , ElasticNet , Random forest , Gradient boosting analizleri uygulanmıştır. Gradient boosting uygun bulunmuştur fakat regresyon analizine çok uyumlu değildir.

LOTUS AI

LOTUS AI

5 MAKİNE ÖRĞENMESİ PROJESİ

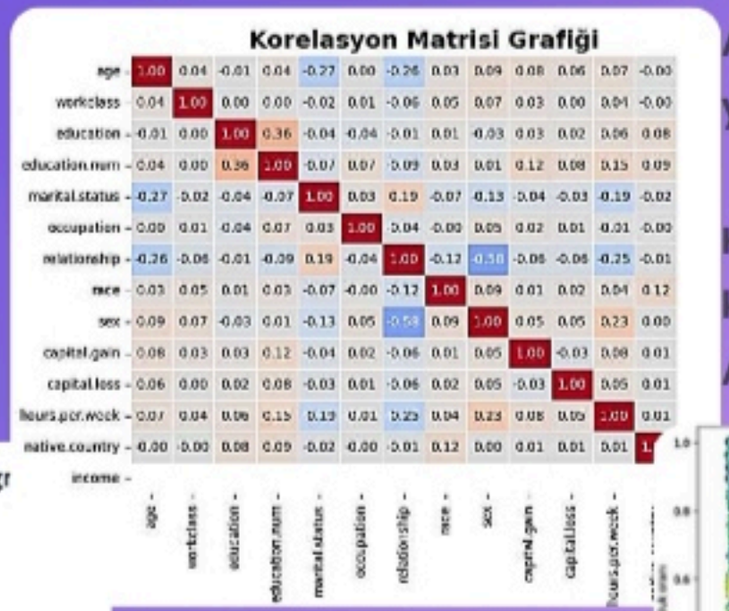
YETİŞKİN NÜFUS SAYIMI GELİRİ ANALİZİ

- **SINIFLANDIRMA ANALİZİ**

```
GridSearchCV kullanıyorum
best_models = {}

for model_name in models:
    print(f"Training {model_name}...")

    model = models[model_name]
    grid = GridSearchCV(estimator=model, param_grid=param_grid[model_name], cv=5, n_jobs=-1, verbose=2
    grid.fit(X_train, y_train)

    best_models[model_name] = grid.best_estimator_

    # en iyi parametre ve sonuç
    print(f"{model_name} - En iyi parametreler: {grid.best_params_}")
    print(f"{model_name} - En iyi skor: {grid.best_score_}")

    y_pred = grid.predict(X_test)

    # başarı değeri
    acc = accuracy_score(y_test, y_pred)
    print(f"{model_name} - Test Accuracy: {acc:.4f}\n")

# tüm sonuçlar
for model_name, model in best_models.items():
    print(f"En iyi model ({model_name}): {model}")
```

```
stackedensemble prediction progr
   predict        p0        p1
0        0  0.991616  0.008384
1        0  0.942138  0.057862
2        0  0.884728  0.115272
3        0  0.924767  0.075233
4        0  0.895883  0.104117
```

- **İLİŞKİ KURALI ANALİZİ**

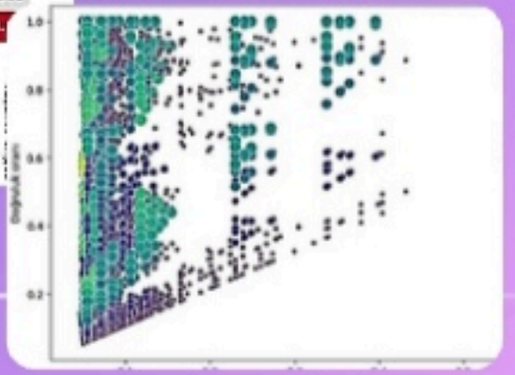


- **SINIFLANDIRMA ANALİZİ**

Random forest , Lojistik regresyon , SVM , KNN , Naive bayes , Decision tree uygulanmıştır. En uygun analiz Lojistik regresyon çıkmıştır. Aynı zamanda AutoML ile de analiz yapılmıştır.

- **İLİŞKİ KURALI ANALİZİ**

Ki-Kare değerleri , Cramer's V değeri , korelasyon matrisi uygulanmıştır. Aynı zamanda AutoMl ile de analiz edilmiştir.



LOTUS AI

LOTUS AI

5 MAKİNE ÖRĞENMESİ PROJESİ

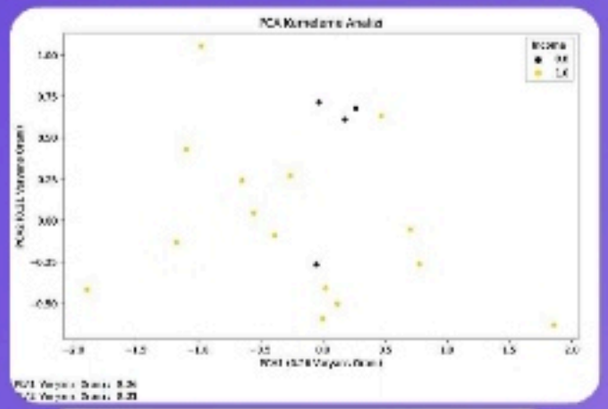
YETİŞKİN NÜFUS SAYIMI GELİRİ ANALİZİ

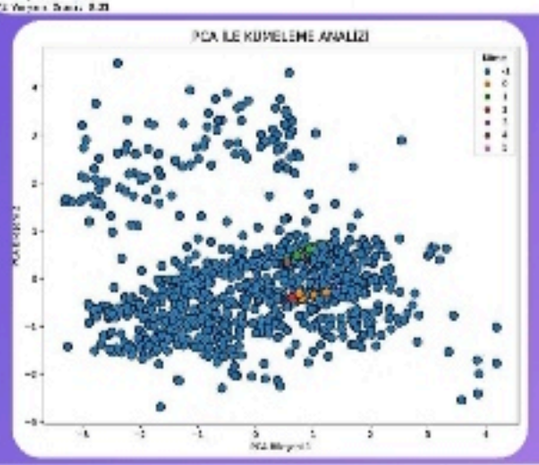
- **KÜMELEME ANALİZİ**

```
from h2o.automl import H2OAutoML

data_h2o['target'] = data_h2o['cluster']

# AutoML eğitimi
aml = H2OAutoML(max_models=10, seed=42, max_runtime_secs=300)
aml.train(x=numerical_cols, y='target', training_frame=data_h2o

aml.leader
```

Cross-Validation Metrics Summary:

	mean	sd	cv_1_valid	cv_2_valid	cv_3_valid	cv_4_
aic	86.49582	58.353813	93.39273	-6.0454764	142.01857	129.6
loglikelihood	0.0	0.0	0.0	0.0	0.0	
mae	0.1348602	0.0189441	0.1365154	0.1029192	0.1531463	0.143
mean_residual_deviance	0.0374231	0.0227488	0.0863113	0.0534249	0.1118028	0.104
mse	0.0374231	0.0227488	0.0863113	0.0534249	0.1118028	0.104
null_deviance	116.70346	14.754734	135.52673	120.26758	110.62369	121.4
r2	0.8467166	0.0450730	0.8695612	0.9124423	0.7969138	0.824
residual_deviance	17.52773	4.7412226	17.40489	10.524704	22.472365	21.2
rmse	0.2884719	0.0402704	0.2937879	0.2311383	0.3343693	0.322
rmsle	0.1717030	0.0217281	0.1669916	0.1379571	0.1864775	0.194





Kümeleme analizinda temel bileşenler analizi (PCA) kullanılmıştır. Sebebi ise varyans oranlarına göre kümeyi en çok temsil eden 2 değişkeni almasıdır. Analiz tamamlanıp yorumlanmıştır.
Aynı zamanda AutoML ile de 5 katlı çapraz doğrulama analizi yapılmış olup sonuçlar değerlendirilmiştir.

LOTUS AI

LOTUS AI

AI AGENT HAKKINDA ARAŞTIRMA RAPORU

- **AI Agent Nedir ?**

AI agent, yapay zeka algoritmalarını kullanarak yapılması gereken görevleri ya da işleri otomatik olarak gerçekleştirebilen ve çevreden aldığı bilgilere göre tepki gösteren bir yazılımdır. Agentler, programlanan talimatlar ve algoritmalar ile hareket ederek karmaşık problemleri çözebilmektedirler. Aynı zamanda kullanıcı ile etkileşimde oldukları için veri analizi de yapabilmektedirler.

- **Temel Özellikleri Nelerdir ?**
- **AI Agent Türleri Nelerdir ?**
- **AI Agentle İle İş Süreçleri Nasıldır ?**

LOTUS AI

TEŞEKKÜR EDERİM

LOTUS AI

YAPAY ZEKA VE BİLİŞİM ÇÖZÜMLERİ A.Ş.