



XGBOOST & STREAMLIT

Ramazan Tetik

- 01** Hakkımda
- 02** XGBoost Nedir?
- 03** Parkinson Projesi
- 04** Sky Data Projesi

SUNUM PLANI



HAKKIMDA

Ben Ramazan Tetik. Aydın Adnan Menderes
Üniversitesi Bilgisayar Mühendisliği
bölümünde 3. sınıfı geçtim. XGBoost ve
Streamlit kullanarak yaptığım projeleri sizlere
sunacağım.

XGBOOST NEDİR?

- XGBoost, denetimli bir topluluk makine öğrenimi algoritmasıdır.
- Gradyan artırma (Gradient Boosting) tekniğine dayalıdır ve her adımda mevcut modelin hatalarını düzeltten yeni ağaçlar ekleyerek modeli güçlendirir.
- XGBoost, bu süreçte hız ve bellek verimliliği sağlayan optimizasyonlara sahiptir ve bu sayede büyük veri setlerinde etkili sonuçlar verir.

XGBOOST ALGORİTMASININ ÖZELLİKLERİ

- **Hızlı ve Etkili:** XGBoost, çok büyük veri kümeleriyle bile hızlı çalışır.
- **Düzenlileştirme teknikleri:** XGBoost, overfitting durumunu önlemeye ve modelin genelleştirme performansını iyileştirmeye yardımcı olan L1 ve L2 düzenlileştirme düzenlileştirme tekniklerini içerir.
- **Eksik değerleri işleme:** Verilerdeki eksik değerleri işlemenin etkili bir yoluna sahiptir.
- **Ağaç budama:** XGBoost, modelin karmaşıklığını azaltmaya ve yorumlanabilirliğini artırmaya yardımcı olan ağaç budamasını destekler.
- **Özellik önemi:** XGBoost, modeldeki her özelliğin önemini ölçmek için bir yol sağlar, bu da özellik seçimi ve modelin davranışını anlamak için yararlı olabilir.

PARKINSON STREAMLIT PROJESİ

Parkinson Disease ↗

Projede, Parkinson hastalığına sahip olan bireyleri sağlıklı bireylerden ayırt etmeye çalışacağız.

Show dataframe

| | PatientID | Age | Gender | Ethnicity | EducationLevel | BMI | Smoking | AlcoholConsumption | Phy |
|---|-----------|-----|--------|-----------|----------------|---------|---------|--------------------|-----|
| 0 | 3,058 | 85 | 0 | 3 | 1 | 19.6199 | 0 | 5.1082 | |
| 1 | 3,059 | 75 | 0 | 0 | 2 | 16.2473 | 1 | 6.0276 | |
| 2 | 3,060 | 70 | 1 | 0 | 0 | 15.3682 | 0 | 2.2421 | |
| 3 | 3,061 | 52 | 0 | 0 | 0 | 15.4546 | 0 | 5.9978 | |
| 4 | 3,062 | 87 | 0 | 0 | 1 | 18.616 | 0 | 9.7752 | |
| 5 | 3,063 | 68 | 1 | 2 | 1 | 39.4233 | 1 | 13.5969 | |
| 6 | 3,064 | 78 | 1 | 0 | 0 | 30.542 | 1 | 2.0113 | |
| 7 | 3,065 | 70 | 1 | 0 | 0 | 36.7583 | 1 | 19.9889 | |
| 8 | 3,066 | 80 | 0 | 2 | 1 | 22.3806 | 1 | 7.2933 | |
| 9 | 3,067 | 71 | 0 | 3 | 2 | 23.7271 | 1 | 17.7829 | |

Description of Data

| | Category | Count |
|---|--------------|-------|
| 0 | Observations | 2,105 |
| 1 | Variables | 34 |
| 2 | cat_cols | 18 |
| 3 | num_cols | 16 |
| 4 | cat_but_car | 0 |
| 5 | num_but_cat | 18 |

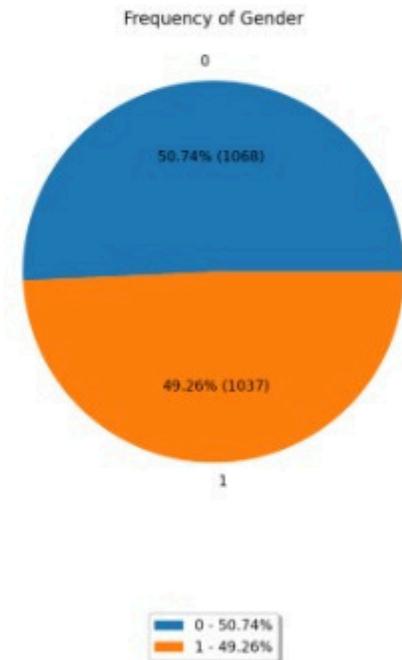
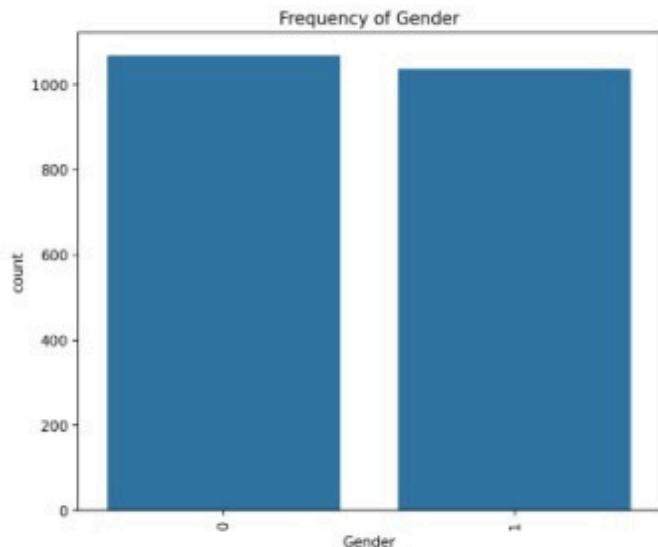
PARKINSON STREAMLIT PROJESİ

Categorical Data Summary

Select a categorical column

Gender

| | Gender | Ratio |
|---|--------|---------|
| 0 | 1,068 | 50.7363 |
| 1 | 1,037 | 49.2637 |

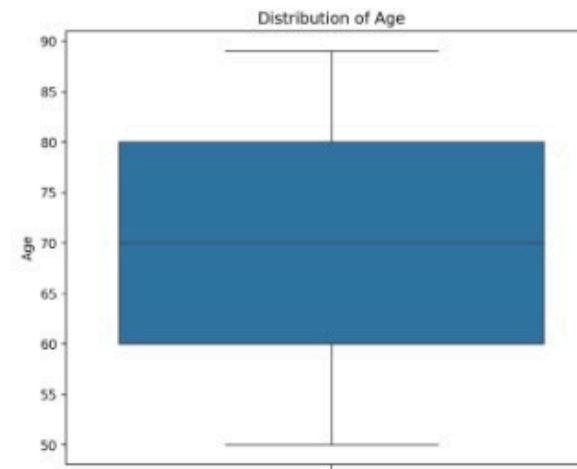
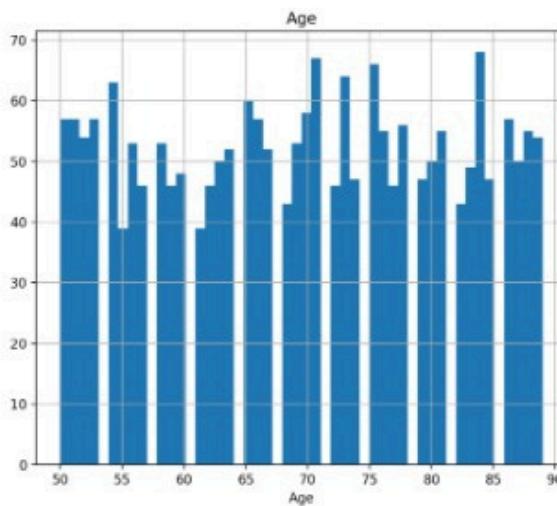


Numerical Data Summary

Select a numerical column

Age

| | Age |
|-------|---------|
| count | 2,105 |
| mean | 69.6019 |
| std | 11.5945 |
| min | 50 |
| 5% | 51 |
| 10% | 53 |
| 20% | 57 |
| 30% | 62 |
| 40% | 66 |
| 50% | 70 |



PARKINSON STREAMLIT PROJESİ

Model Evaluation Metrics

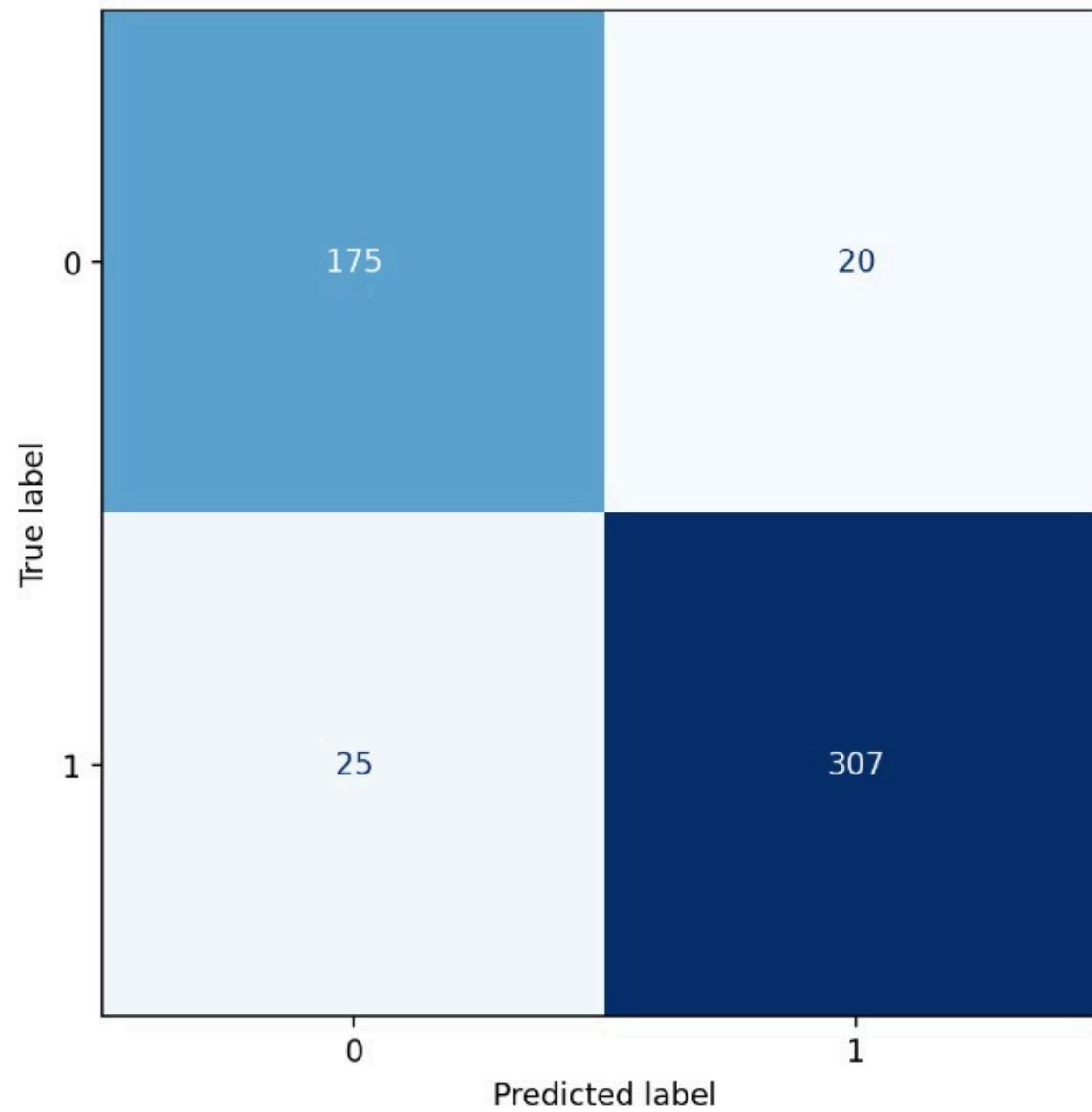
Model Performance Metrics:

| | Model | Accuracy | Recall | Precision | F1 Score | AUC |
|---|------------------------|----------|--------|-----------|----------|------|
| 0 | LogisticRegression | 0.79 | 0.855 | 0.82 | 0.84 | 0.77 |
| 1 | KNeighborsClassifier | 0.68 | 0.771 | 0.73 | 0.75 | 0.64 |
| 2 | DecisionTreeClassifier | 0.89 | 0.913 | 0.92 | 0.92 | 0.89 |
| 3 | RandomForestClassifier | 0.91 | 0.925 | 0.94 | 0.93 | 0.91 |
| 4 | SVC | 0.78 | 0.849 | 0.81 | 0.83 | 0.76 |
| 5 | XGBClassifier | 0.91 | 0.925 | 0.94 | 0.93 | 0.91 |

Select a model

XGBClassifier

Confusion Matrix for XGBClassifier:



SKY DATA PROJESİ

Sky Data Classification and Visualization

This project explores classification techniques applied to a dataset containing observations of stars, galaxies, and quasars.

Dataset Overview ↴

Show dataframe

| | ra | dec | u | g | r | i | z | class | redshift | plate | mjd |
|---|----------|--------|---------|---------|---------|---------|---------|--------|----------|-------|-------|
| 0 | 183.5313 | 0.0897 | 19.4741 | 17.0424 | 15.947 | 15.5034 | 15.2253 | STAR | 0 | 3,306 | 54,92 |
| 1 | 183.5984 | 0.1353 | 18.6628 | 17.2145 | 16.6764 | 16.4892 | 16.3915 | STAR | -0.0001 | 323 | 51,61 |
| 2 | 183.6802 | 0.1262 | 19.383 | 18.1917 | 17.4743 | 17.0873 | 16.8013 | GALAXY | 0.1231 | 287 | 52,02 |
| 3 | 183.8705 | 0.0499 | 17.7654 | 16.6027 | 16.1612 | 15.9823 | 15.9044 | STAR | -0.0001 | 3,306 | 54,92 |
| 4 | 183.8833 | 0.1026 | 17.5503 | 16.2634 | 16.4387 | 16.5549 | 16.6133 | STAR | 0.0006 | 3,306 | 54,92 |
| 5 | 183.8472 | 0.1737 | 19.4313 | 18.4678 | 18.1645 | 18.0148 | 18.0416 | STAR | 0.0003 | 324 | 51,66 |
| 6 | 183.8644 | 0.0192 | 19.3832 | 17.89 | 17.1054 | 16.6639 | 16.3696 | GALAXY | 0.1002 | 287 | 52,02 |
| 7 | 183.9001 | 0.1875 | 18.9799 | 17.845 | 17.3802 | 17.2067 | 17.0707 | STAR | 0.0003 | 3,306 | 54,92 |
| 8 | 183.9246 | 0.0972 | 17.9062 | 16.9717 | 16.6754 | 16.5378 | 16.476 | STAR | 0.0001 | 323 | 51,61 |
| 9 | 183.9735 | 0.0816 | 18.6725 | 17.7138 | 17.4936 | 17.2828 | 17.2264 | GALAXY | 0.0405 | 288 | 52,00 |

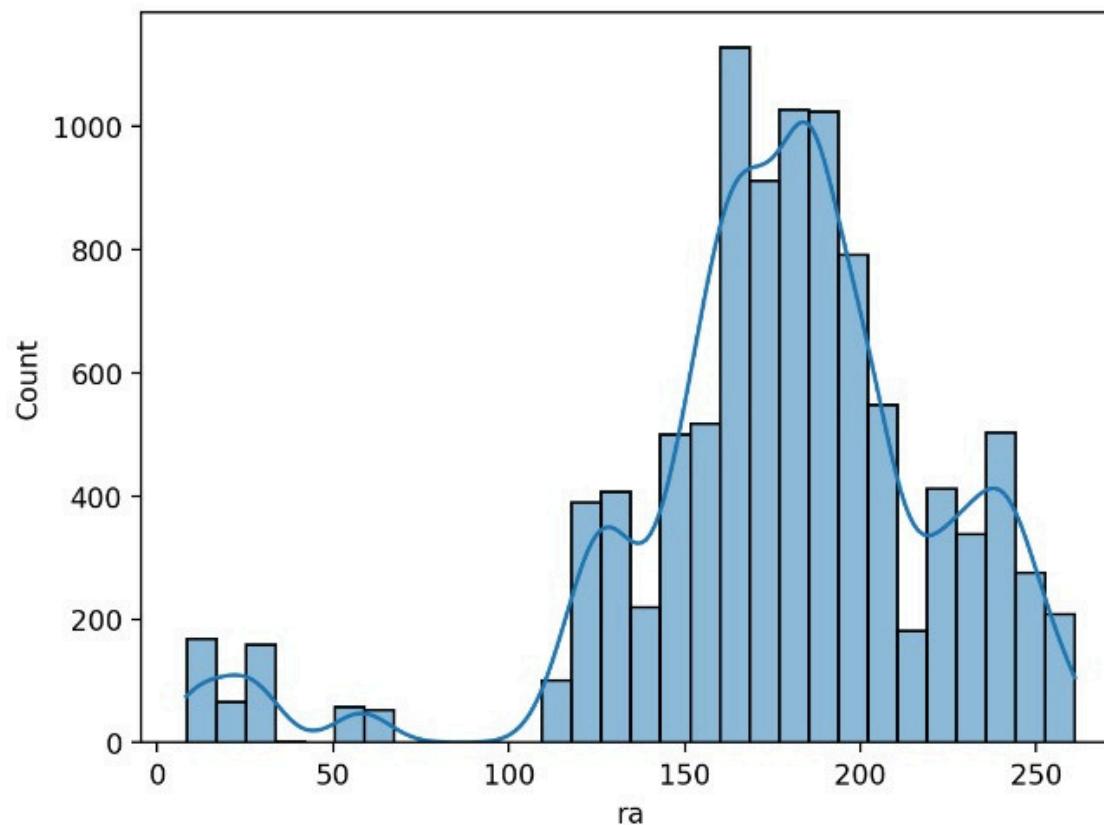
SKY DATA PROJESİ

Numerical Features Visualization

Select numerical feature for visualization:

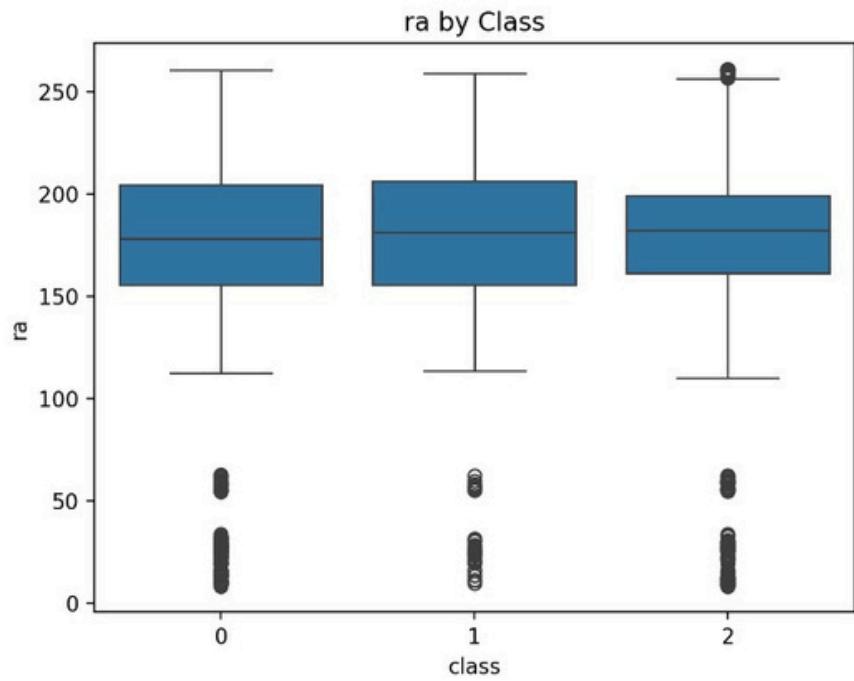
ra

Distribution of ra

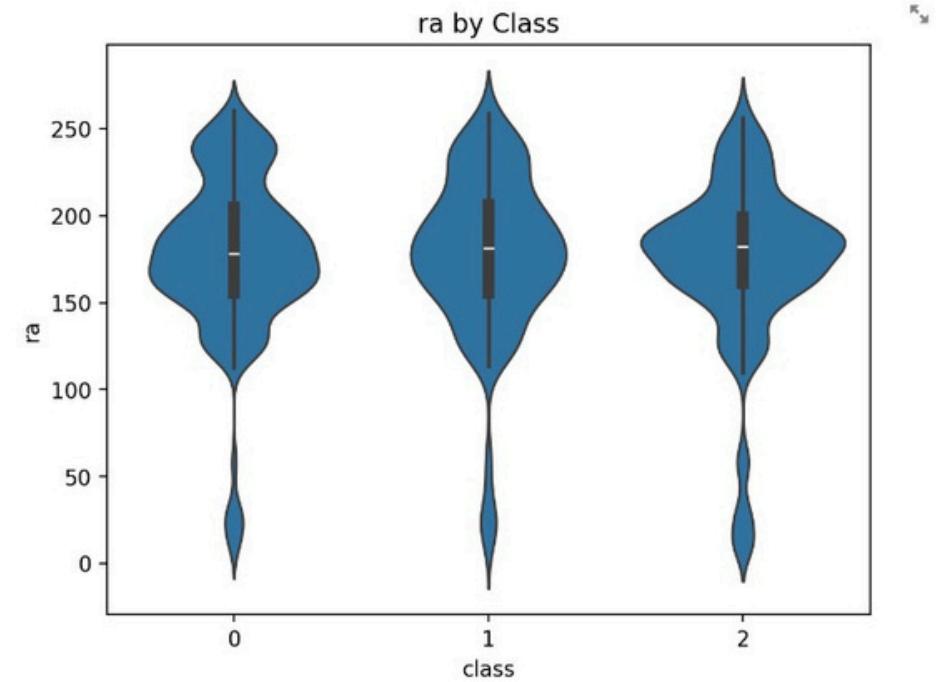


SKY DATA PROJESİ

Box Plot of ra



Violin Plot of ra



SKY DATA PROJESİ

Model Results

| | Model | Accuracy (%) | Training Time (s) | Prediction Time (s) |
|---|----------------------|--------------|-------------------|---------------------|
| 0 | K-Nearest Neighbors | 92.7576 | 0.0118 | 0.7100 |
| 1 | Gaussian Naive Bayes | 98.0000 | 0.0379 | 0.0025 |
| 2 | Random Forest | 99.1515 | 3.6570 | 0.0467 |
| 3 | XGBoost | 99.1818 | 0.9062 | 0.0403 |

Model Comparison

| | Model | Accuracy (%) |
|---|----------------------|--------------|
| 0 | K-Nearest Neighbors | 92.7576 |
| 1 | Gaussian Naive Bayes | 98.0000 |
| 2 | Random Forest | 99.1515 |
| 3 | XGBoost | 99.1818 |

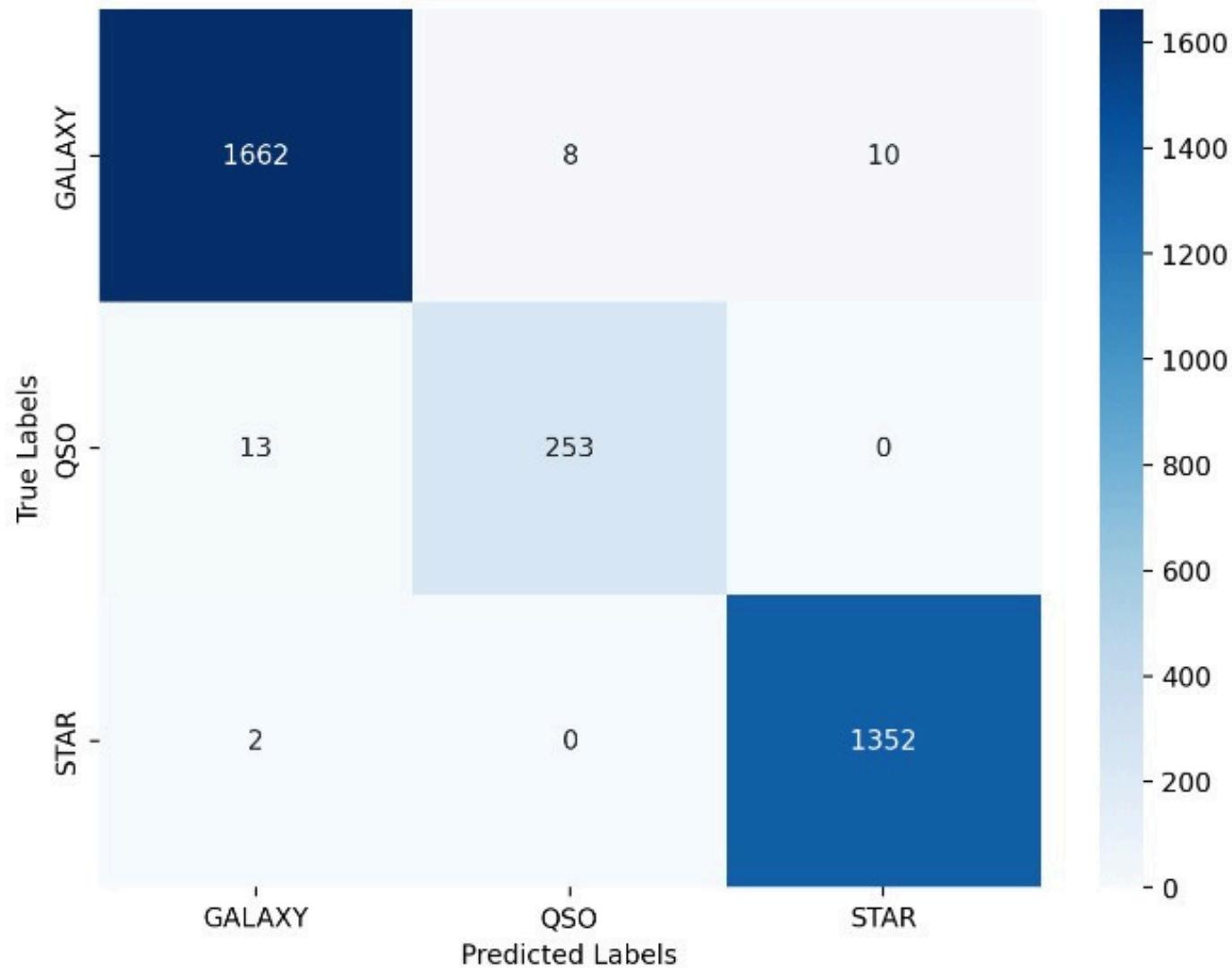
Select a model to view the confusion matrix:

XGBoost



XGBoost Confusion Matrix

XGBoost Confusion Matrix



TEŞEKKÜRLER