



TRAKYA ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ
BLM421 PROJE II RAPORU

PROJE ADI:

EEG Sinyallerinden Psikolojik Rahatsızlık Analizi

PROJE EKİBİ:

Beyda Bucak – 1211602099

Ahmet Sezer Darıcı - 1201602026

PROJE DANIŞMANI:

Dr. Öğr. Üyesi Deniz Mertkan GEZGİN

Edirne - 2025

İÇİNDEKİLER

1. GİRİŞ & PROJENİN TANITIMI-----	3
2. GEREKSİNİM ANALİZİ -----	4
2.1. Çalışma Platformları-----	4
2.2. Donanım Gereksinimi -----	4
2.3. Kütüphaneler -----	4
2.4. Geliştirme Ortamı-----	4
2.5. Çözüm Bulunan Sorunlar -----	4
3. TASARIM -----	5
3.1. Genel Mimari-----	5
3.2. Kullanıcılar İçin Sonuç-----	5
4. KODLARLA AÇIKLAMA -----	6
4.1. Veri Yükleme ve İnceleme -----	7
4.2. Korelasyon ve Özellik Seçimi-----	8
5. SONUÇ-----	11
5.1. Gelecekte Yapılabilecek Geliştirmeler-----	7

1. GİRİŞ & PROJENİN TANITIMI

Bu projede, EEG (Elektroensefalografi) sinyalleri üzerinden bireylerin psikolojik sağlık durumlarının analiz edilmesi amaçlanmıştır. Özellikle delta ve theta beyin dalgalarının, sağlıklı bireylerle psikolojik rahatsızlık (örneğin şizofreni) tanısı konmuş bireyler arasında anlamlı farklılıklar gösterdiği tespit edilmiştir.

Veri analiz süreci Jamovi yazılımı üzerinden başlatılmış, ardından EEG sinyalleri .csv formatında dışa aktarılmış ve Python ortamında Jupyter Lab kullanılarak veri temizliği ve makine öğrenmesi süreçleri gerçekleştirilmiştir.

Çeşitli sınıflandırma modelleri test edilerek en yüksek doğruluğu veren **Gradient Boosting Classifier** modeli seçilmiştir. Bu model, delta ve theta dalgalarına göre bireyin psikolojik durumunu % olarak tahmin etmektedir. Son aşamada, bu modeli kullanarak Flask tabanlı bir web arayüzü geliştirilmiş ve HTML/CSS ile kullanıcı dostu bir web sayfası hazırlanmıştır.

2. GEREKSİNİM ANALİZİ

2.1. Çalışma Platformları

- Python 3.x
- Jamovi (statistiksel analiz için)
- Jupyter Lab
- Flask
- HTML / CSS

2.2. Donanım Gereksinimi

- En az 4GB RAM
- Intel i5 veya eşdeğeri işlemci

2.3. Kütüphaneler

- pandas
- numpy
- scikit-learn
- matplotlib
- seaborn
- flask

2.4. Geliştirme Ortamı

- Jupyter Notebook
- Online HTML IDE

2.5. Çözüm Bulunan Sorunlar

- Eksik veriler temizlendi
- Veri ölçekleme ile model başarımı artırıldı

- Veri dengesizliği oversampling ile giderildi
- En uygun model seçimi yapılarak doğruluk maksimize edildi

3. TASARIM

Proje aşağıdaki aşamalardan oluşmaktadır:

3.1. Veri Kümesinin Yüklenmesi

Jamovi üzerinden analiz edilen EEG sinyalleri, CSV formatında dışa aktarılmış ve Python ortamına yüklenmiştir.

3.2. Veri Ön İşleme

Eksik veriler kaldırılmış, etiketler sayısal hale getirilmiş ve delta/theta özellikleri çıkarılmıştır. Veriler ölçeklendirilmiş ve dengesizlik sorunu veri artırma ile çözülmüştür.

3.3. Modelleme

Aşağıdaki sınıflandırma modelleri test edilmiştir:

- Logistic Regression
- Random Forest
- Gradient Boosting
- KNN
- SVM

En başarılı model **Gradient Boosting Classifier** olmuştur.

3.4. Model Değerlendirme

Model, **doğruluk (accuracy)**, **confusion matrix**, **learning curve** gibi metriklerle değerlendirilmiştir.

3.5. Arayüz ve Tahmin

Flask ile web arayüzü oluşturulmuş ve HTML/CSS ile kullanıcı giriş ekranı geliştirilmiştir. Kullanıcı delta ve theta değerlerini girerek sistemden tahmin alabilmektedir.

3.6. Kullanıcılar İçin Sonuç

Geliştirilen sistem, EEG delta ve theta dalgalarını giren bir kullanıcıya, psikolojik olarak sağlıklı mı yoksa bir bozukluk riski mi taşıdığına dair yüzde tahmini sunmaktadır. Sistem gelecekte hastanelerde ön tanı aracı olarak veya mobil uygulama olarak geliştirilebilir.

4. KODLARLA AÇIKLAMA

4.1. Veri Yükleme ve İnceleme

	no.	sex	age	eeg.date	education	IQ	main.dis...	specific...	AB.A.delt...	AB.A.delt...
1	1	M	57.00	2012.8.30			Addictive dis...	Alcohol use ...	35.999	21.717
2	2	M	37.00	2012.9.6	6	120	Addictive dis...	Alcohol use ...	13.425	11.003
3	3	M	32.00	2012.9.10	16	113	Addictive dis...	Alcohol use ...	29.942	27.545
4	4	M	35.00	2012.10.8	18	126	Addictive dis...	Alcohol use ...	21.496	21.847
5	5	M	36.00	2012.10.18	16	112	Addictive dis...	Alcohol use ...	37.776	33.608
6	6	F	24.00	2012.11.21	14	105	Addictive dis...	Alcohol use ...	13.482	14.096
7	7	F	26.00	2012.12.3	16	103	Addictive dis...	Alcohol use ...	21.781	26.655
8	8	M	23.00	2013.1.17	12	104	Addictive dis...	Alcohol use ...	11.704	10.600
9	9	M	24.00	2013.2.15	17	89	Addictive dis...	Alcohol use ...	8.532	8.668
10	10	F	30.00	2013.2.8	12	98	Addictive dis...	Alcohol use ...	14.854	12.614
11	11	M	34.00	2013.2.18	13	91	Addictive dis...	Alcohol use ...	37.149	43.506
12	12	M	32.00	2013.3.8	12	93	Addictive dis...	Alcohol use ...	26.104	26.800
13	13	M	20.00	2013.2.20	13	105	Addictive dis...	Alcohol use ...	30.399	32.379
14	14	M	28.00	2013.4.15	14	118	Addictive dis...	Alcohol use ...	15.840	29.317
15	15	M	23.00	2013.4.11	13	116	Addictive dis...	Alcohol use ...	16.085	15.888
16	16	M	20.00	2013.4.25	12	93	Addictive dis...	Alcohol use ...	16.620	14.605
17	17	M	26.00	2013.8.29	12	107	Addictive dis...	Alcohol use ...	16.338	17.716
18	18	M	30.00	2013.9.27		86	Addictive dis...	Alcohol use ...	12.443	12.504
19	19	M	31.00	2013.10.18	18	102	Addictive dis...	Alcohol use ...	14.806	14.013
20	20	M	25.00	2013.11.20	16	120	Addictive dis...	Alcohol use ...	34.557	27.539
21	21	M	39.00	2013.12.27	12	128	Addictive dis...	Alcohol use ...	18.595	17.021
22	22	M	20.00	2014.10.23		116	Addictive dis...	Alcohol use ...	28.287	22.412
23	23	M	30.00	2014.7.7	16	102	Addictive dis...	Alcohol use ...	31.039	22.107
24	24	M	23.00	2014.7.25	12	104	Addictive dis...	Alcohol use ...	18.802	17.190
25	25	M	29.00	2014.7.25	13	122	Addictive dis...	Alcohol use ...	28.266	23.889
26	26	M	32.00	2014.9.19	16	130	Addictive dis...	Alcohol use ...	23.862	25.012
27	27	M	34.00	2014.9.25	13	122	Addictive dis...	Alcohol use ...	35.108	42.888
28	28	M	26.00	2014.12.18	17	108	Addictive dis...	Alcohol use ...	16.084	19.081

Jamovi platformu üstünden incelenen bu verisetinde yaklaşık 1146 adet sütun, 946 adet denek bulunmaktadır. Deneye her katılan bireyin yaş, cinsiyet, toplam aldığı eğitim yılı gibi temel bilgileri girilmiştir. Eksik veriler python ile temizlenmiş veya uygun parametrelerle doldurulmuştur.

4.2. Korelasyon ve Özellik Seçimi

Independent Samples T-Test

Independent Samples T-Test

		Statistic	p
AB_alpha	Mann-Whitney U	10486	0.348
AB_beta	Mann-Whitney U	9420	0.084
AB_gamma	Mann-Whitney U	11274	0.723
COH_delta	Mann-Whitney U	10884	0.521
COH_theta	Mann-Whitney U	10900	0.528
COH_alpha	Mann-Whitney U	9719	0.132
COH_beta	Mann-Whitney U	11496	0.850
COH_gamma	Mann-Whitney U	11312	0.745
AB_delta	Mann-Whitney U	8200	0.009
AB_theta	Mann-Whitney U	8841	0.031

Note. $H_a \mu_2 \neq \mu_1$

Yapılan analizde delta ve theta dalgalarının sağlıklı ve sağlıklı olmayan bireyler arasında anlamlı bir fark yarattığı görülmüştür ($p < 0.05$).

4.3. Model Eğitimi ve Değerlendirme

```
X = df_balanced[features]
y = df_balanced[target]
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y, test_size=0.2, random_state=42)
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

models = {
    "Logistic Regression": LogisticRegression(max_iter=1000),
    "SVM": SVC(kernel='rbf', probability=True),
    "KNN": KNeighborsClassifier(n_neighbors=5),
    "Random Forest": RandomForestClassifier(n_estimators=100, max_depth=6, min_samples_split=10, min_samples_leaf=5,
                                           class_weight='balanced', random_state=42),
    "Gradient Boosting": GradientBoostingClassifier(n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42)
}

results = []

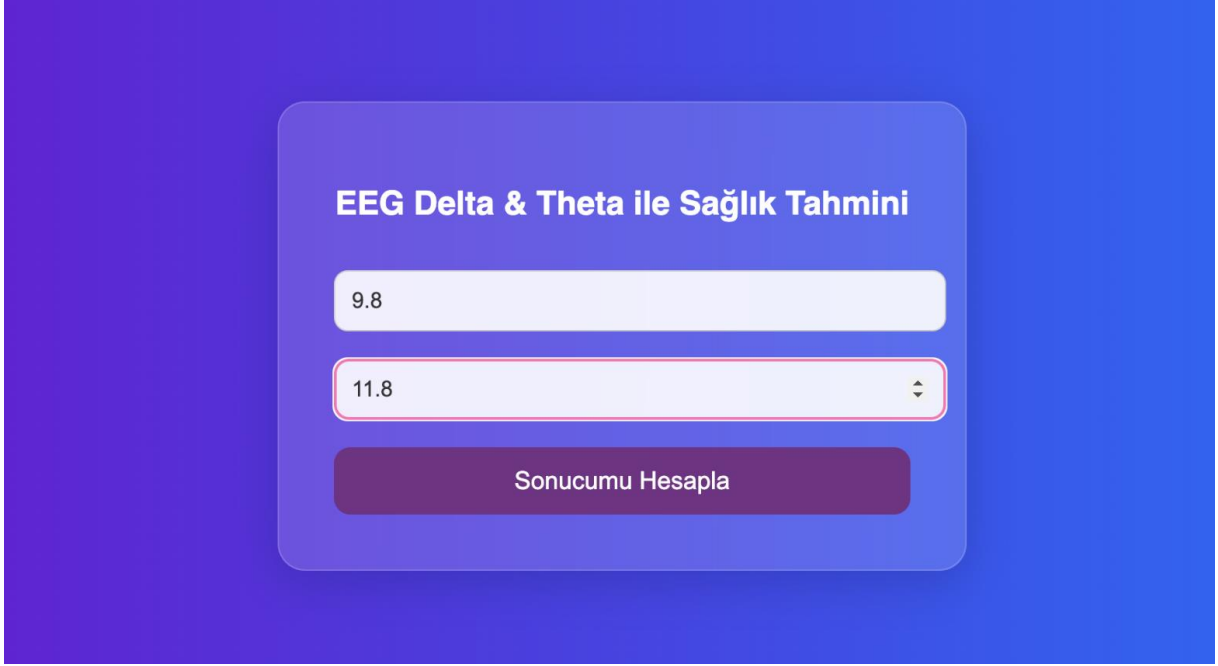
for name, model in models.items():
    model.fit(X_train_scaled, y_train)
    y_pred = model.predict(X_test_scaled)
    acc = accuracy_score(y_test, y_pred)
    results.append((name, acc))
    print(f"\n {name} Accuracy: {acc:.4f}")
    print(classification_report(y_test, y_pred, target_names=["Healthy", "Disorder"]))

best_model = max(results, key=lambda x: x[1])
print(f"\n En iyi model: {best_model[0]} ({best_model[1]:.4f} doğruluk)")
```


Logistic Regression Accuracy: 0.6581				
	precision	recall	f1-score	support
Healthy	0.00	0.00	0.00	91
Disorder	0.66	0.99	0.79	181
accuracy			0.66	272
macro avg	0.33	0.49	0.40	272
weighted avg	0.44	0.66	0.53	272
SVM Accuracy: 0.7096				
	precision	recall	f1-score	support
Healthy	0.68	0.25	0.37	91
Disorder	0.71	0.94	0.81	181
accuracy			0.71	272
macro avg	0.70	0.60	0.59	272
weighted avg	0.70	0.71	0.66	272
KNN Accuracy: 0.9265				
	precision	recall	f1-score	support
Healthy	0.82	1.00	0.90	91
Disorder	1.00	0.89	0.94	181
accuracy			0.93	272
macro avg	0.91	0.94	0.92	272
weighted avg	0.94	0.93	0.93	272
Random Forest Accuracy: 0.9265				
	precision	recall	f1-score	support
Healthy	0.82	1.00	0.90	91
Disorder	1.00	0.89	0.94	181
accuracy			0.93	272
macro avg	0.91	0.94	0.92	272
weighted avg	0.94	0.93	0.93	272
Gradient Boosting Accuracy: 0.9449				
	precision	recall	f1-score	support
Healthy	0.93	0.90	0.92	91
Disorder	0.95	0.97	0.96	181
accuracy			0.94	272
macro avg	0.94	0.93	0.94	272
weighted avg	0.94	0.94	0.94	272
En iyi model: Gradient Boosting (0.9449 doğruluk)				

Yukarıda görülen model sonuçlarına göre, Gradient Boosting en fazla doğruluğa, en az hata oranına sahip modeldir. Ayrıca 1.0 skoru hiçbir parametrede gözükmemektedir. Bu, modelin overfit olmadığına işaret etmektedir.

4.4. Görselleştirme ve Arayüz



The screenshot shows a web application interface with a blue gradient background. In the center, there is a white rounded rectangle containing the title "EEG Delta & Theta ile Sağlık Tahmini". Below the title, there are two input fields. The first input field contains the value "9.8". The second input field contains the value "11.8" and has a small downward arrow icon on its right side. Below these input fields, there is a dark blue button with the text "Sonucumu Hesapla" in white.

Flask kullanarak yapılan websitesinde, theta ve delta sinyallerinin sayısal değerinin girildiği takdirde sağlıklı/sağlıksız durumunu ekrana yansıtır, bir olasılık belirler ve pie chart üzerinde gösterir. Spektogram için örnek bir görüntü grafiği oluşturarak kullanıcıya kolay ve anlaşılabilir bir arayüz sunar.

5. SONUÇ

Bu proje kapsamında EEG sinyalleri analiz edilerek psikolojik rahatsızlık tahmini üzerine bir sınıflandırma sistemi geliştirilmiştir. Delta ve theta dalgalarının anlamlı ayırım gücü olduğu görülmüş ve Gradient Boosting modeliyle yüksek doğruluk elde edilmiştir.

Sistemin sunduğu faydalar şunlardır:

- **Doğruluk:** Verimli veri temizliği ve model seçimi sayesinde yüksek başarı sağlanmıştır.
- **Uygulanabilirlik:** Web arayüzü sayesinde kullanıcıya kolay erişim sunulmuştur.
- **Genelleme:** Model, test verisinde de başarılı sonuçlar vererek overfitting göstermemiştir.

5.1. Gelecekte Yapılabilecek Geliştirmeler

- EEG frekans bantlarına alfa, beta gibi yeni dalgalar eklenebilir
- Daha büyük veri setleriyle model yeniden eğitilebilir
- Web arayüzü mobil uyumlu hale getirilebilir
- Derin öğrenme algoritmaları ile performans karşılaştırması yapılabilir

KAYNAKÇA

Jamovi Statistical Software. <https://www.jamovi.org/> (Eriřim Tarihi: 10 Nisan 2025)

Scikit-learn Documentation. <https://scikit-learn.org/> (Eriřim Tarihi: 10 Nisan 2025)

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (2nd Ed.). O'Reilly Media.

Flask Web Framework. <https://flask.palletsprojects.com/>

Python Software Foundation. <https://www.python.org/>