

1. Temel Beklenti ve Özgürlük

- Amaç:** Kendi belirlediğiniz bir sektördeki probleme, **uçtan uca makine öğrenmesi** ile çözüm geliştirmek ve **her adımı dökümant etmek**.
- Önemli:** Sektörü, problemi ve datayı **siz seçeceksiniz**.
- Zorluk Seviyesi:** Projenin zorluğu, **seçeceğiniz dataset** ve **vereceğiniz emekle** doğru orantılıdır. Gözünüz korkmasın, repo karmaşık görünse de adım adım ilerleyebilirsınız.

2. Dataset Seçimi ve Kriterler (⚠️ Çok Önemli)

Dataseti **Kaggle'daki geçmiş veya devam eden yarışmalardan** bulmanızı şiddetle tavsiye ediyoruz.

Zorunlu Kriterler:

- Format:** **Tabular (Tablo) data** olmalı (.csv, .parquet, .xlsx). Görüntü veya ses **olmamalı**.
- Gerçeklik:** **Asla sentetik (yapay) veri olmamalı**. Kaggle "Playground" serisi genelde sentetiktir, **kullanmayın**.
- Boyut:** En az **10.000 satır** veri olmalı. (Feature Engineering yapabilmek için).
- Öznitelik:** Tercihen **80-90 kolon (feature)** olması iyidir, ancak en az 10-15 kolonun altına düşmeyin.

3. Proje Yapısı ve Adımlar

Projenin Kaggle yarışmasından farkı; sadece yüksek skor değil, **Business kurgusu ve Sistemsel Düzen** istememizdir.

Notebook Akışı Şöyleden Olmalıdır:

- EDA:** Veriyi kendiniz yorumlayın. (LLM çıktısını kopyalamayın, **analitik yorum katın**).
- Baseline Model:** En basit modelle bir referans skoru elde edin.
- Feature Engineering (Kritik):** Yaratıcılığınızı konuşturun. Dışarıdan veri ekleyin (altın, faiz vb.).
- Model Optimization:** Hiperparametre optimizasyonu (Optuna, GridSearch vb.).
- Model Evaluation:** Feature importance, SHAP analizi ve **Business yorumu**.
- Deployment (Arayüz):** Modeli **Streamlit** veya **Flask** ile basit bir arayüze bağlayın. (Hugging Face Spaces veya Render üzerinde çalışabilir).

4. Repoda Cevaplanması ZORUNLU Sorular

Jüri repo'ya baktığında şu 8 sorunun cevabını net görmeli:

- Problem Tanımı:** Neyi çözüyorsun?
- Baseline Skoru:** İlk basit modelin başarısı neydi?
- Feature Engineering:** Hangi değişkenleri türettin? (**Yaratıcılık burada ölçülür**).
- Validasyon Şeması:** K-Fold mu, Time-based mi? **Neden bunu seçtin?** (Yanlış seçim projeyi çöp edebilir).
- Ön İşleme Stratejisi:** Final feature setine nasıl karar verdin?
- Final vs Baseline Farkı:** Ne kadar iyileştirdin ve fark nereden geliyor?

7. **Business Uyumu:** Model iş gereksinimlerini karşılıyor mu?
8. **Monitoring (İzleme):** Modeli canlıya alsak **Data Drift'i (veri kaymasını)** nasıl izlerdik?

5. "Olsa Güzel Olur" (Bonus Puanlar

Projeyi standarttan **profesyonel seviyeye** taşıyan maddeler:

- **Git Geçmiş:** "Upload files" değil, adım adım atılmış **commitler**.
- **Monitoring Sistemi:** Tahminlerin loglandığı basit bir veritabanı/ekran.
- **Business Kurgusu:** Kendinizi bir şirkette hayal edip çizdiğiniz **akış şeması**.
- **Üst Yönetim Sunumu:** Teknik olmayan bir yöneticiye (C-Level) projeyi satan **5-10 slaytlık sunum**.
- **YouTube Videosu & Medium Yazısı:** Projenin anlatımı.

6. Önemli Uyarılar & Tavsiyeler

- **Kod Temizliği (Clean Code):** Notebook'unuzu **emojilerle, HTML süslemeleriyle** veya ChatGPT'den kopyalanmış **gerekisiz uzun yorumlarla** doldurmayın. Kodunuz **sade ve okunabilir** olsun.
- **Derin Öğrenme (Deep Learning):** Kullanabilirsiniz ama önce **Machine Learning (XGBoost, LightGBM, RF)** ile temel atın.
- **İletişim:** Takıldığınız yerde WhatsApp grubundan sorun. Birbirinizin rakibi değilsiniz.
- **Son Teslim Tarihi: 9 Aralık** (Yaklaşık 3 hafta süre var).

Kaynaklar

- **Örnek Repo Yapısı:** [Enes Manan - Credit Risk Model](#)
- **Proje Rehberi:** [Made With ML](#)
- **Mülakat Kitabı:** [ML Engineering Book \(Burkov\)](#)

[Data Science Awesome Repo](#)

Zero2End ML Bootcamp final projesi, katılımcıların kendi seçikleri bir sektör ve probleme yönelik, uçtan uca bir makine öğrenmesi çözümü geliştirmelerini hedeflemektedir. Bu süreçte katılımcılardan Kaggle gibi platformlardaki gerçekçi yarışma verilerini kullanmaları, buna karşın sentetik verilerden ve 'Playground' serisi yarışmalardan kesinlikle kaçınmaları beklenmektedir. Projenin temel amacı sadece yüksek bir model skoru elde etmek değil; 10.000 satır ve üzeri tabular (tablo) formatındaki verilerle çalışarak, gerçek bir iş problemini (Business Case) ele alan sistemsel ve bütüncül bir kurgu oluşturmaktır.

Teknik uygulama aşamasında, düzenli bir GitHub repo yapısı (docs, models, notebooks, src) ve belirli bir notebook akışı (EDA, Baseline, Feature Engineering, Optimizasyon, Evaluation, Pipeline) zorunludur. Katılımcıların özellikle Feature Engineering adımında yaratıcılıklarını sergilemeleri, seçikleri validasyon şemalarını (K-Fold, Time-based vb.) neden seçiklerini mantıklı gerekçelerle açıklamaları gerekmektedir. Projenin finalinde, eğitilen modelin Streamlit gibi araçlarla basit bir arayüze bağlanarak canlıya alınması (Deployment) ve README dosyasında problem tanımı, baseline-final farkı ve iş gereksinimlerine uyum gibi kritik soruların net bir şekilde cevaplanması şarttır.

Projenin profesyonel bir portfolyo niteliği taşıması için düzenli bir Git geçmişi oluşturulması ve Clean Code prensiplerine uyularak gereksiz emoji veya HTML süslemelerinden kaçınılması önemle vurgulanmıştır. Ayrıca, projeyi bir adım öne taşımak isteyenler için üst yönetim sunumu, YouTube videosu veya Medium yazısı gibi bonus içerikler önerilmektedir. 9 Aralık son teslim tarihine kadar katılımcıların birbirleriyle iletişim halinde kalarak yardımlaşmaları, ezberci yaklaşımından uzak durarak analitik yorumlama yeteneklerini sergilemeleri ve projeyi yaşayan bir ürün gibi tasarlamaları beklenmektedir.