

Bitirme projemde, ilk olarak proje hedefleri ve gereksinimleri belirledim. Bu adımı tamamladıktan sonra, projenin yapımında kullanılacak araç ve teknolojileri seçtim. Ardından veri toplama ve işleme sürecini planladım, veri setinin nereden alınacağına karar verdim ve veri setini bulup local bilgisayarıma indirdim.

Sonrasında, projenin ana hatlarını anlatan ve gidişat hakkında yol gösterici olmasını beklediğim İş Zaman Planını çıkardım.

FAZ	GÖREVLER/HAFTALAR	FEB		MAR				APR				MAY				JUN			
		15	22	1	8	15	22	1	8	15	22	1	8	15	22	1	8	15	22
1	Planlama ve Hazırlık	Proje hedefleri ve gereksinimlerinin belirlenmesi																	
		Proje araçları ve teknolojilerinin seçimi																	
		Veri toplama ve işleme sürecinin planlanması																	
		Bitirme projesi hakkında başlık ve genel özet yazılması ve volkan hocaya teslim edilmesi																	
		İş zaman planının çıkarılması																	
2	Veri İşleme ve Analiz	Docker Engine kurulumunun yapılması																	
		Veri seti işlemleri için Spark ve Python ortamının kurulması																	
		Veri setinin analiz edilmesi																	
		Veri ön işleme ve temizleme işlemlerinin yapılması																	
		Veri görselleştirmelerinin yapılması																	
3	Model Geliştirme ve Eğitim	Model eğitimi ve Test için kullanılacak veri setlerinin ayrıştırılması																	
		PySpark MLib literatür araştırması ve kaynak taramasının yapılması																	
		Veri setinin model eğitimi için hazırlanması																	
		Makine öğrenmesi modelinin seçimi ve geliştirilmesi																	
		Modelin eğitilmesi, optimum parametrelerin bulunması																	
4	Gerçek Zamanlı Uygulama ve Test	Model performansının değerlendirilmesi																	
		Kafka Elasticsearch ve Kibana kurulumlarının yapılması																	
		Test setinin Kafka producer ile simüle edilmesi																	
		Simüle edilen verinin Real Time olarak Elasticsearch'te depolanması																	
		Elasticsearch'te depolanan verinin Real Time olarak kibana'da görselleştirilmesi																	
		Simüle edilen Real Time verinin Spark Streaming ile yakalanması ve ML modelinin uygulanması																	
		Model çıktılarının ayrı ayrı Kafka Topiclerine gönderilmesi																	
		Model çıktılarının Elasticsearch'te depolanıp Kibana'da görselleştirilmesi ve Alert mekanizmasının Oluşturulması																	
		Projenin adımlarının scriptleştirilmesi																	
		Sistem performansının değerlendirilmesi																	

İlk olarak VirtualBox üzerinde geliştirme yapmayı planlamıştım fakat çok daha rahat ve hızlı bir ortam sunduğu için Docker Engine'e geçiş yapmaya karar verdim ve kurulumlarını yaptım.

PySpark ortamını Docker'a kurarak veri seti işlemleri gerçekleştirdim. Veri setini analiz ederek, ön işleme ve temizleme işlemlerini yaptıktan sonra, verileri görselleştirmeye başladım. Verilerin görselleştirilmesiyle, proje hakkında daha iyi bir fikir edindim ve verilerin daha iyi anlaşılmasını sağladım. Son olarak da, model eğitimi ve test için kullanılacak veri setlerini ayrıştırdım.

Bu adımların hepsini iş zaman planı doğrultusunda, planlı bir şekilde, gerçekleştirdim.

Projenin sonraki aşamalarında, PySpark MLib literatür araştırması ve kaynak taraması yapılacak ve veri seti model eğitimi için hazırlanacaktır. Bu adımda, uygun makine öğrenmesi modeli seçilerek geliştirilecektir. Model eğitimi gerçekleştirilerek, optimum parametreler bulunacaktır. Ayrıca, modelin performansı da değerlendirilecektir.

Daha sonra, Kafka, Elasticsearch ve Kibana kurulumları da Docker üzerinde gerçekleştirilecek ve test seti Kafka Producer ile simüle edilecektir. Simüle edilen veriler gerçek zamanlı olarak Elasticsearch'te depolanacak ve depolanan veriler gerçek zamanlı olarak Kibana'da görselleştirilecektir. (Elasticsearch'ün Anomaly Detection özelliğinin de projeye dahil edilmesi hedeflenmektedir. Bunun için bir Google Cloud hesabı gerekmekte olduğu için ilerleyen aşamalarda karar verilecektir.)

Simüle edilen gerçek zamanlı veriler Spark Streaming ile yakalanacak ve makine öğrenmesi modeli uygulanacaktır. Model çıktıları ayrı ayrı Kafka Topiclerine gönderilerek bir Alert mekanizması oluşturulacaktır.

Bu aşamaların gerçekleştirilmesiyle birlikte, projenin son halinin tamamlanması ve çalıştırılabilir hale getirilmesi hedeflenmektedir.