

BLM3760
UZMAN SİSTEMLER
DÖNEM PROJESİ

16011047
Kutay YILMAZ




16011056
Özcan KARABACAK

CRIMES IN CHICAGO

Konu: Chicago’ da işlenen suçların detaylarının bulunduğu veri setini kullanarak seçilen durumlarda karşılaşılabilecekleri suçların karar ağacı kullanılarak çıkarımının yapılması.

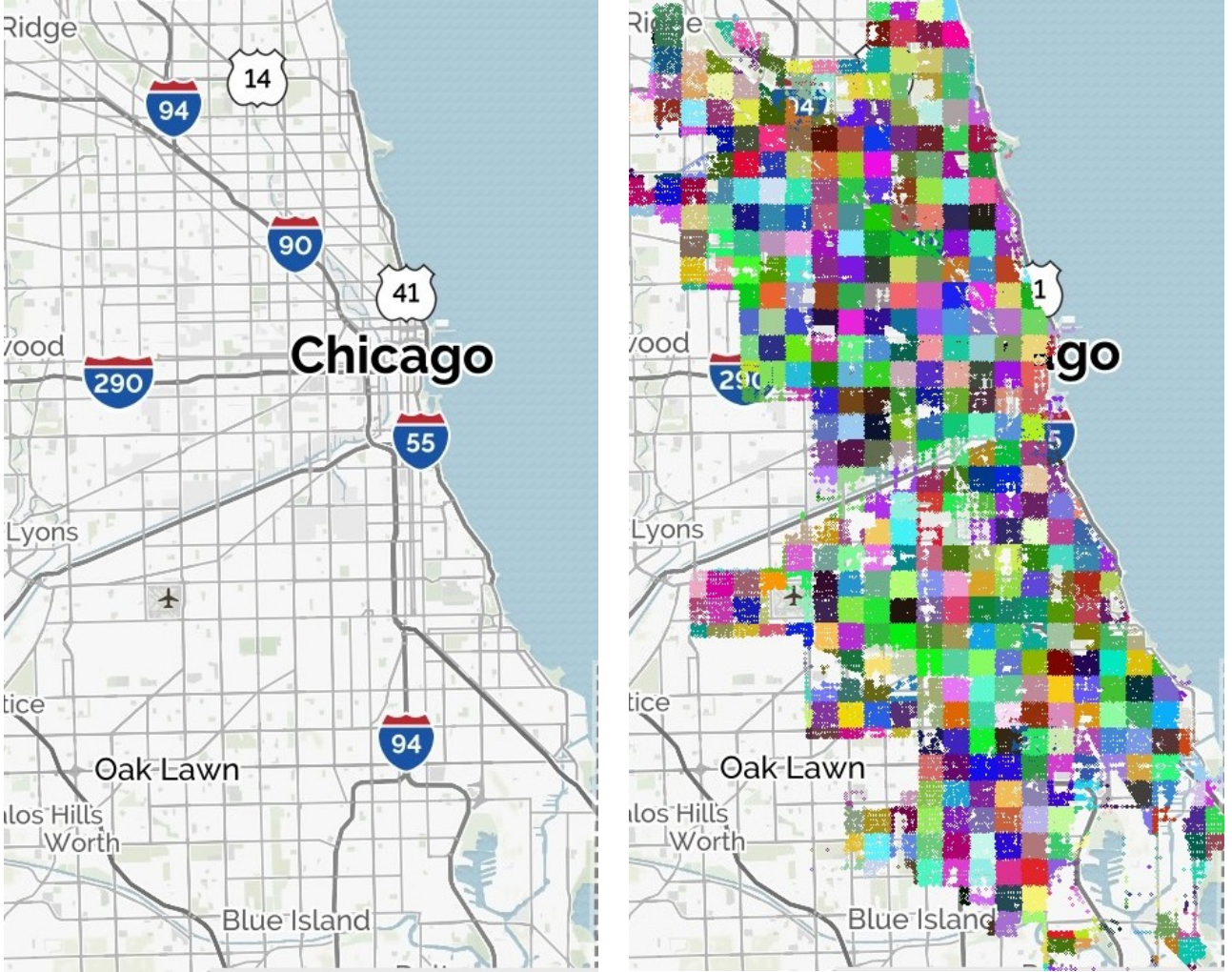
Karar ağacının veri setinde olmayan verilerin ağaçtaki soru ilişkilerine göre değerlendirilebilmesi, aranacak değerin ağaç yapısı sayesinde logaritmik karmaşıklıkta bulunabilmesi gibi özellikleri klasik arama yöntemlerinin sağlayamadığı özellikler arasındadır.

Veri setindeki hangi sütunların kullanılmasının daha mantıklı çıkarımlar yapacağı tartışıldı. Saat, suç tipi, enlem, boylam ve lokasyon verilerinin kullanılması uygun görüldü. Kullanılan verilerin bir kısmı aşağıda gösterilmiştir.

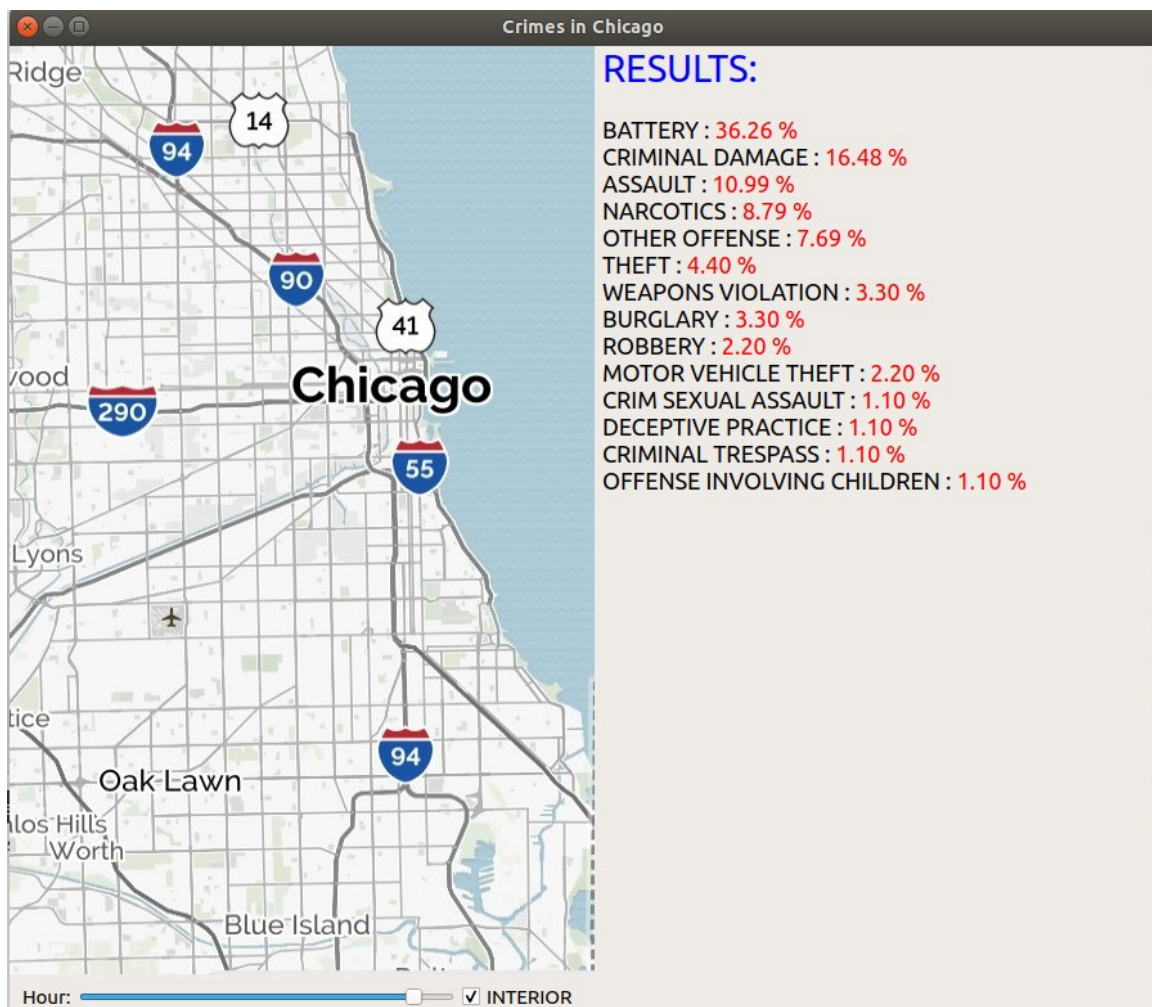
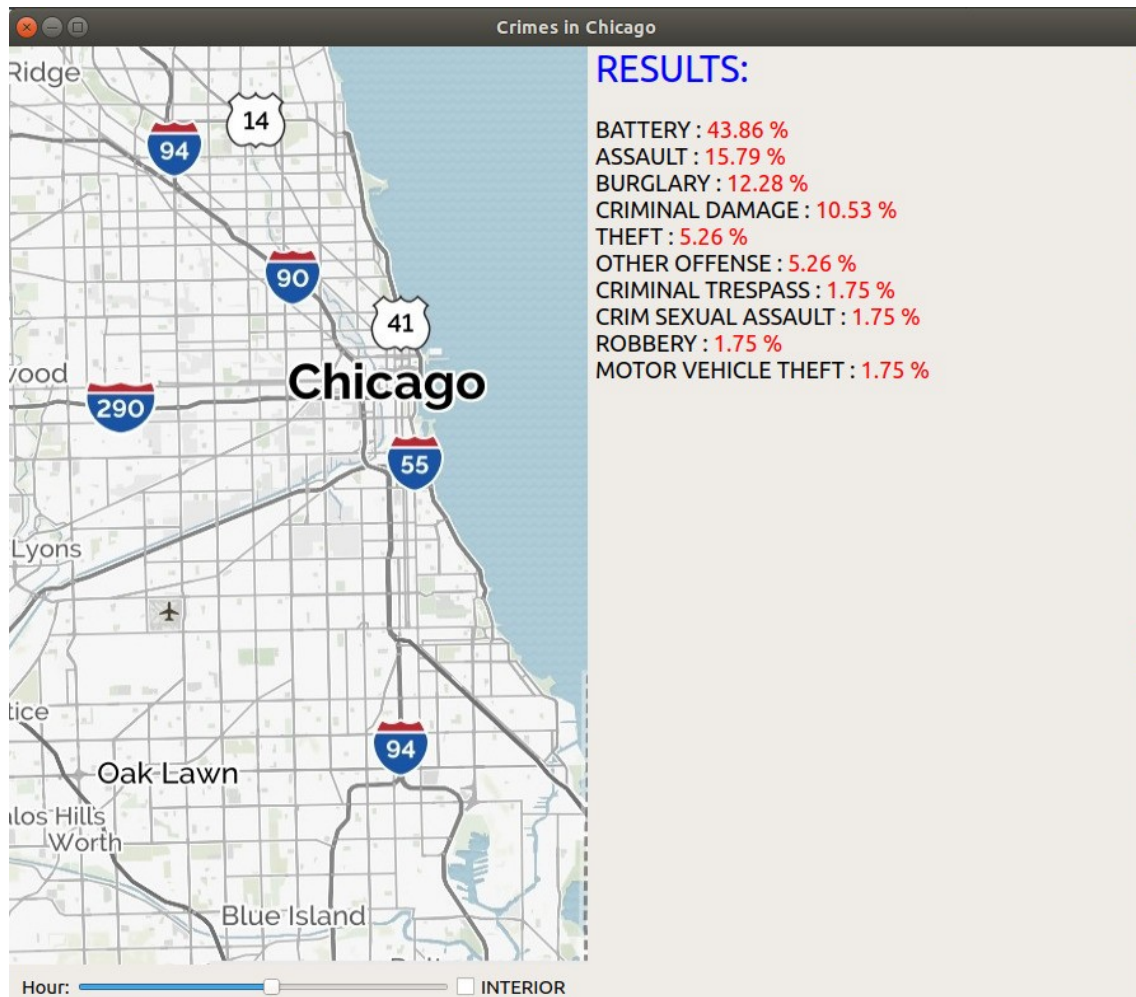
	 Date	 Primary Type	 Location Description	# Latitude	# Longitude
1	05/03/2016 11:40:00 PM	BATTERY	APARTMENT	41.864073157	-87.706818608
2	05/03/2016 09:40:00 PM	BATTERY	RESIDENCE	41.782921527	-87.60436317
3	05/03/2016 11:31:00 PM	PUBLIC PEACE VIOLATION	STREET	41.894908283	-87.758371958
4	05/03/2016 10:10:00 PM	BATTERY	SIDEWALK	41.885686845	-87.749515983
5	05/03/2016 10:00:00 PM	THEFT	RESIDENCE	41.886297242	-87.761750709
6	05/03/2016 10:35:00 PM	BATTERY	STREET	41.745354023	-87.603798903
7	05/03/2016 10:30:00 PM	BATTERY	CHA HALLWAY/STAIRWELL/EL EVATOR	41.844023772	-87.626923253
8	05/03/2016 09:30:00 PM	BATTERY	RESIDENCE PORCH/HALLWAY	41.811133958	-87.62074077
9	05/03/2016 04:00:00 PM	BATTERY	SIDEWALK	41.99813061	-87.665814038
10	05/03/2016 10:30:00 PM	BATTERY	STREET	41.768096835	-87.663878589
11	05/03/2016 09:44:00 PM	WEAPONS VIOLATION	VEHICLE NON-COMMERCIAL	41.761733286	-87.558309979
12	05/03/2016 11:11:00 PM	BATTERY	SIDEWALK	41.893026751	-87.626750829

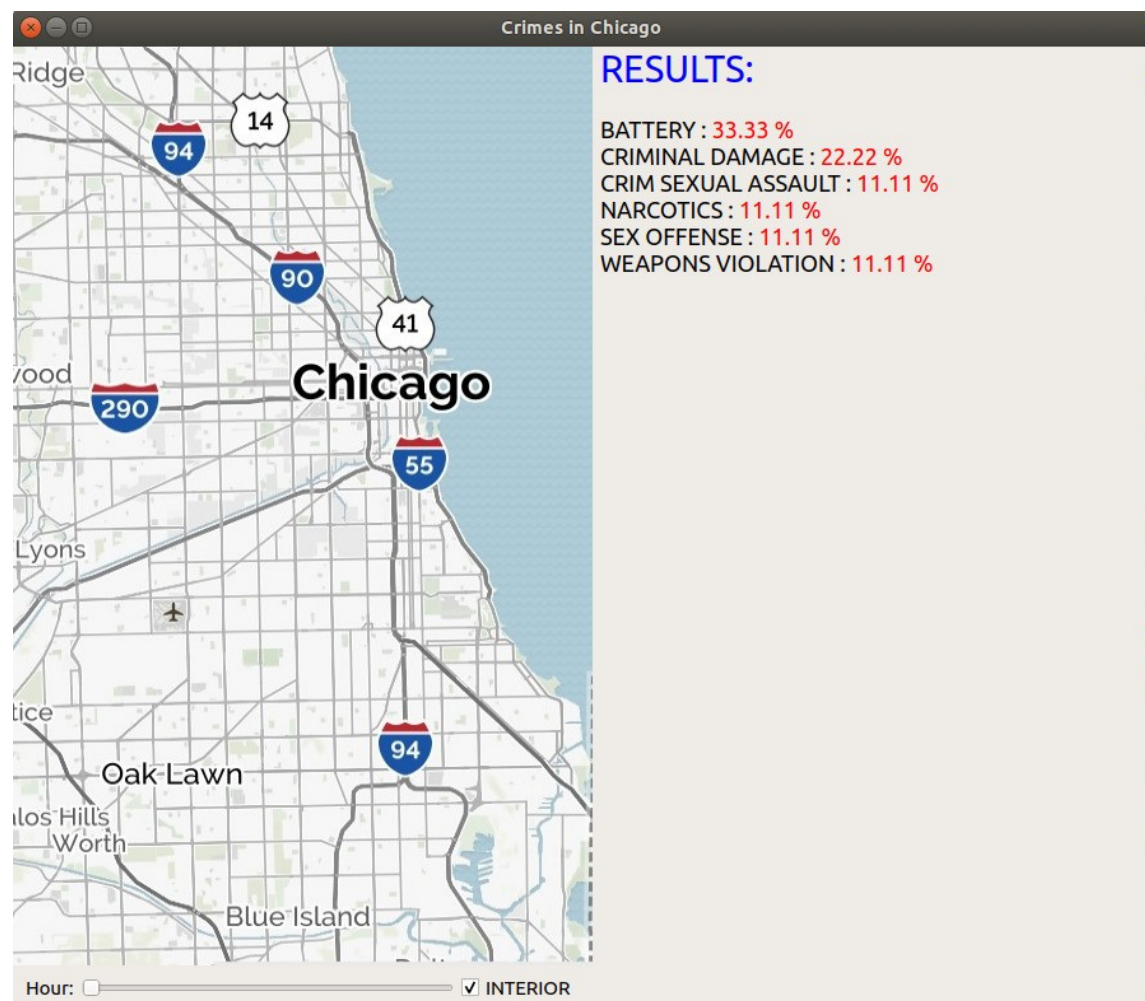
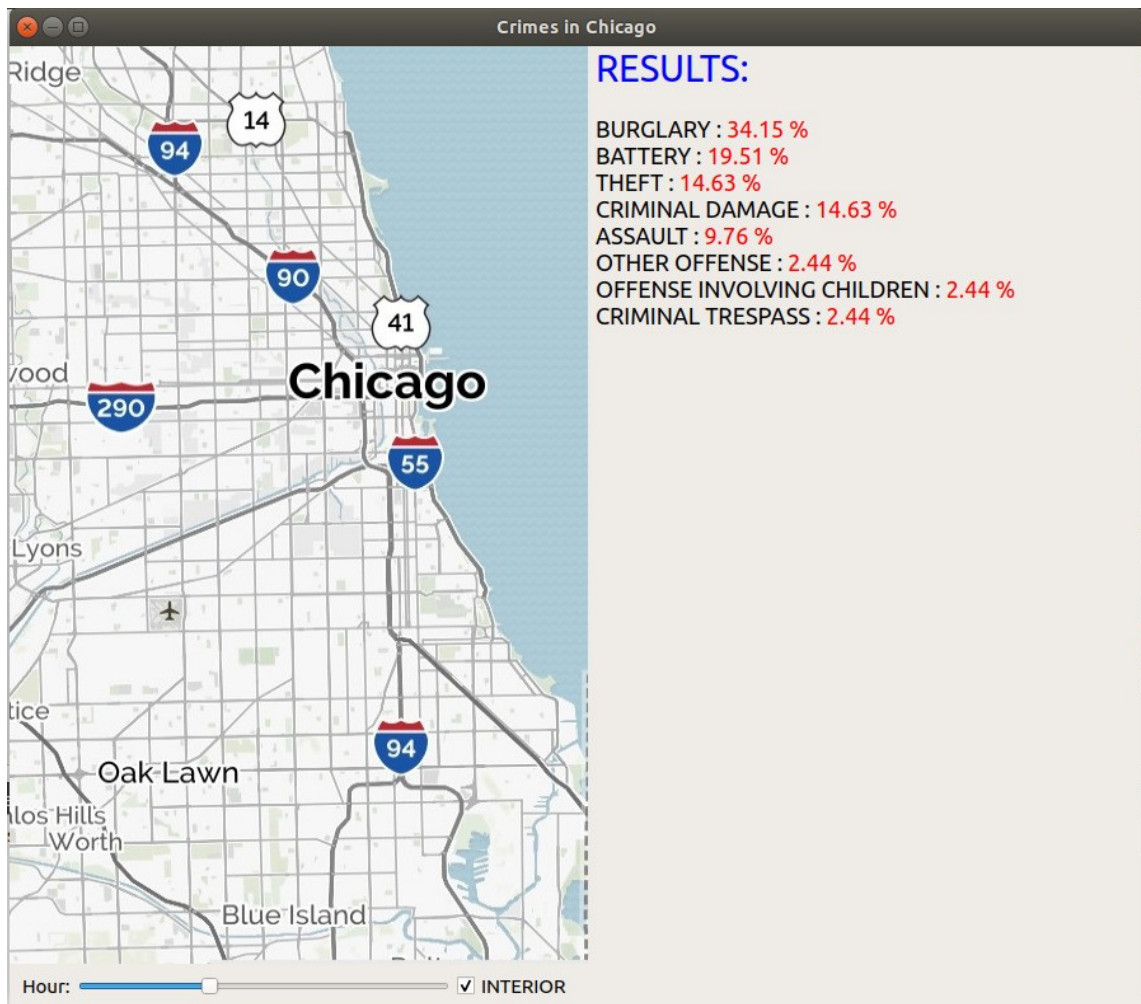
Bu verilerde gerekli düzenlemeler yapıldı. Enlem ve boylam bilgileri kullanılarak suçlar bölgelere ayrıldı. Tarih kısmında yalnızca saat bilgisinin kullanılmasının daha uygun olacağına karar verildi. Lokasyon bilgileri de iç ortam ve dış ortam olacak şekilde düzenlendi.

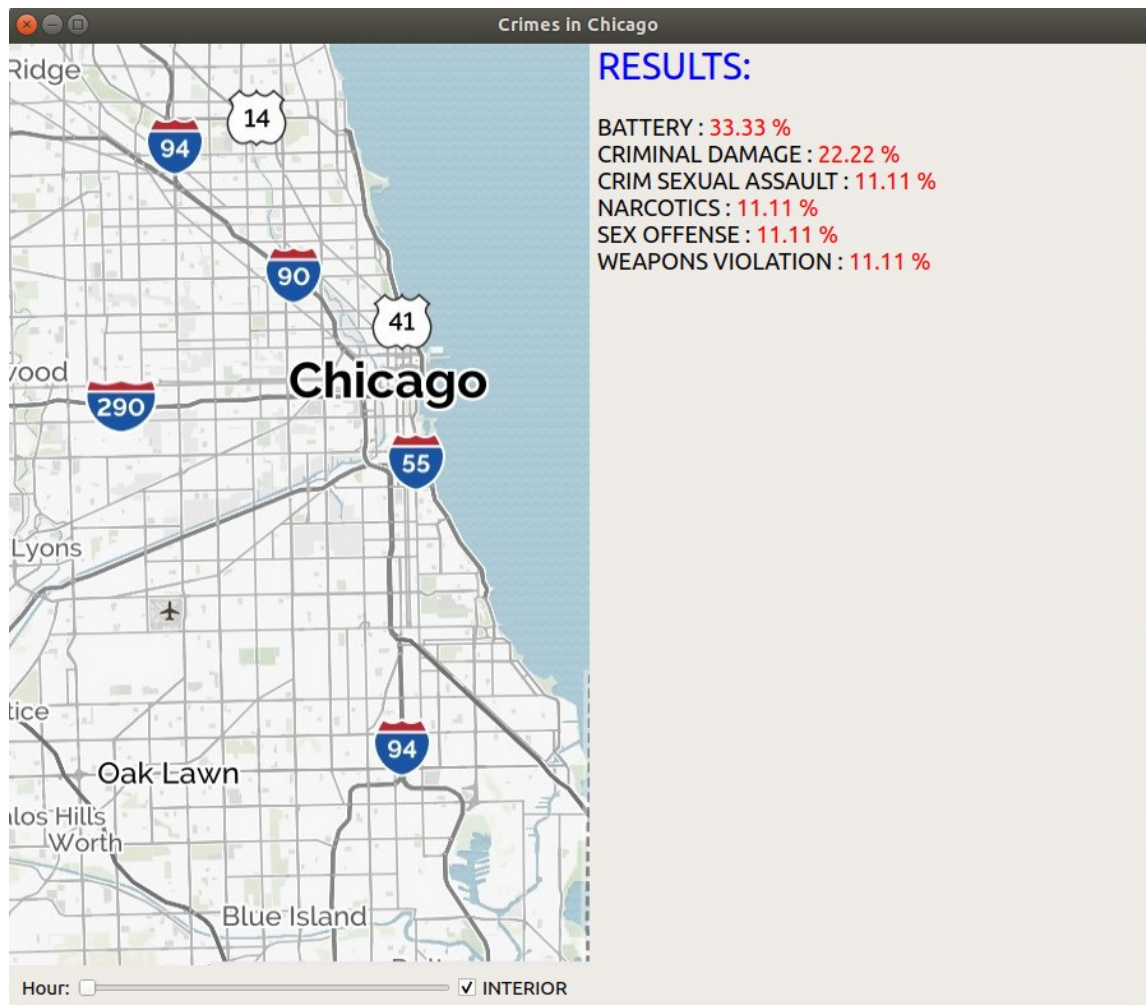
2012-2017 yılları arası işlenen suçlar harita üzerinde görselleştirildi.



Karar ağacı oluşturulurken bir CART algoritması olan Gini Impurity yöntemi kullanıldı. GUI kısmında ise kullanıcıya konum, saat ve ortam bilgisi sorulmaktadır. Seçim sonucunda, girilen parametreler karar ağacında değerlendirilerek kullanıcıya karşılaşılabileceği suçların oranları gösterilmektedir. Aşağıda programın örnek çıktıları verilmiştir.







Geliştirme Süreci:

Nesne yönelimli programlama metodolojisi kullanılmıştır. C++ ve Qt ortamıyla geliştirilmiştir. Karar ağacının implementasyonu internetten araştırıldı. Veri seti düzenlenirken ve harita üzerinde suçların görselleştirilmesi için Python dili kullanılmıştır.

SİSTEMİN SAYISAL BAŞARISI

Başarının Ölçülmesi:

Rastgele karıştırılmış veri setinin yaklaşık %80' lik (1.000.000 veri) kısmı karar ağacını oluşturmak için, kalan %20 lik (226.000 veri) kısmı başarıyı test etmek için kullanmıştır.

Başarının Hesaplanması:

Test verilerinden üretilen sonuçların test verisinin etiketini içerdiği durumların yüzdesi ve bu başarılı tahminlerin ortalama yüzde kaç oranla kullanıcıya önerildiği hesaplanmıştır.

Başarı Oranı:

Test verisinin etiketi %93.851 oranla tahmin verilerinin içerisinde bulunmaktadır. Ortalama %19.473 oranla tehlike etiketlenmiştir.

Değerlendirme:

Yüksek oranda karşılaşılabilecek tehlike bilinebilmiştir ve sistem asıl etiketin tahminlerdeki yüzdesinin yaklaşık %20 olması sebebiyle ortalama 5 tehlike tahmininde kullanıcıyı doğru tehlikeye karşı uyarabilir olduğunu göstermiştir.

KAYNAK

<https://www.kaggle.com/currie32/crimes-in-chicago>

https://github.com/random-forests/tutorials/blob/master/decision_tree.py

<https://www.qt.io/>

https://en.wikipedia.org/wiki/Decision_tree