

The most important features for this model are the offense descriptions and offense code.

Offense Descriptions refers to the detailed narrative or definition of a criminal act. An offense description provides clarity about what the crime involves, helping law enforcement officials, legal professionals, and the public to understand the nature of the crime. For example, the offense description for burglary might specify it as the unlawful entry into a building with the intent to commit theft or any other felony.

Offense Codes are just numeric or alphanumeric codes assigned to specific types of criminal offenses. They serve as a shorthand method to categorize and record offenses in legal and criminal databases. Offense codes simplify reporting and statistical analysis, aiding in data management and research within criminal justice systems. For example, different jurisdictions might have unique codes for theft, assault, or drug possession, which can be used in police reports and court documents.

District code also is one of the most important features even though it is considered as a protected feature. District code is the location where the crime happened. AI systems can inadvertently learn and perpetuate geographic biases present in the training data. For example, a credit scoring AI might deny loans to people from certain neighborhoods more frequently if historical data shows higher default rates in those areas. This can lead to discrimination against individuals based solely on where they live, which might correlate with racial, economic, or social factors. Location often correlates with various socio-economic indicators such as income, education level, and access to services. AI models that use location data might inadvertently reinforce existing inequalities if not carefully managed. By treating location as a protected feature, developers can work to ensure their models do not deepen social divides.

Other protected features such as street, latitude, and longitude was already removed back in the data cleaning section due the ethical considerations. Ethical AI practices demand that algorithms do not create or exacerbate unfair treatment of individuals based on their geographical location. This includes ensuring equal access to opportunities and benefits provided by AI technologies, regardless of an individual's location.

Random 5 random samples are chosen from the testing dataset with a fixed random state for reproducibility.

	OFFENSE_CODE	OFFENSE_DESCRIPTION	DISTRICT	OCCURRED_ON_DATE	MONTH	DAY_OF_WEEK	HOUR	Severe_crimes
508	3207	85	8	1	1	1	8	0
969	3802	74	4	3	1	3	18	0
377	613	51	8	0	1	0	10	0
2788	801	6	3	14	1	0	12	1
2083	3201	86	10	10	1	3	14	0

The random forest model trained from week 10 is used to predict on these samples and get the result

```
# use the model to make predictions
predictions = model.predict(test_sample)
predictions

[8] ✓ 0.0s
of .. array([0, 0, 0, 1, 0], dtype=int64)
```

If we change the Offense code to 801 of the 4<sup>th</sup> sample, the result will be inverted no matter the description of the crime. Also, If we change the 5<sup>th</sup> sample's District to 1, it will also be inverted to severe crime. The model is biased and consider certain district with a set of time will turn the crime to severe state.

To remove this bias and also stop using the protected features, the column of district code is removed from both train, test and validation dataset. A new random forest model is built and trained using the new cleaned data. If this process is not done and law enforcement are using the legacy model. It will cause the majority number of resources being used in few certain districts.