

Module 5 Project

Wildfire Causes Prediction



By Mukesh Kumar Chaudhary

Outline

- About Project
- Data Gathering
- Data Exploration and Analysis
- Modelling and Result Analysis
- Conclusion

ABOUT PROJECT

- Predicting causes of wildfire .
- Is arson cause main cause in future ?



DATA

- Sqlite
- Fires table
- 1880465 rows * 39 columns
- 8 features for ML
- Balance Subset 188047 rows
- STAT_CAUSE_DESCR
- FIRE SIZE

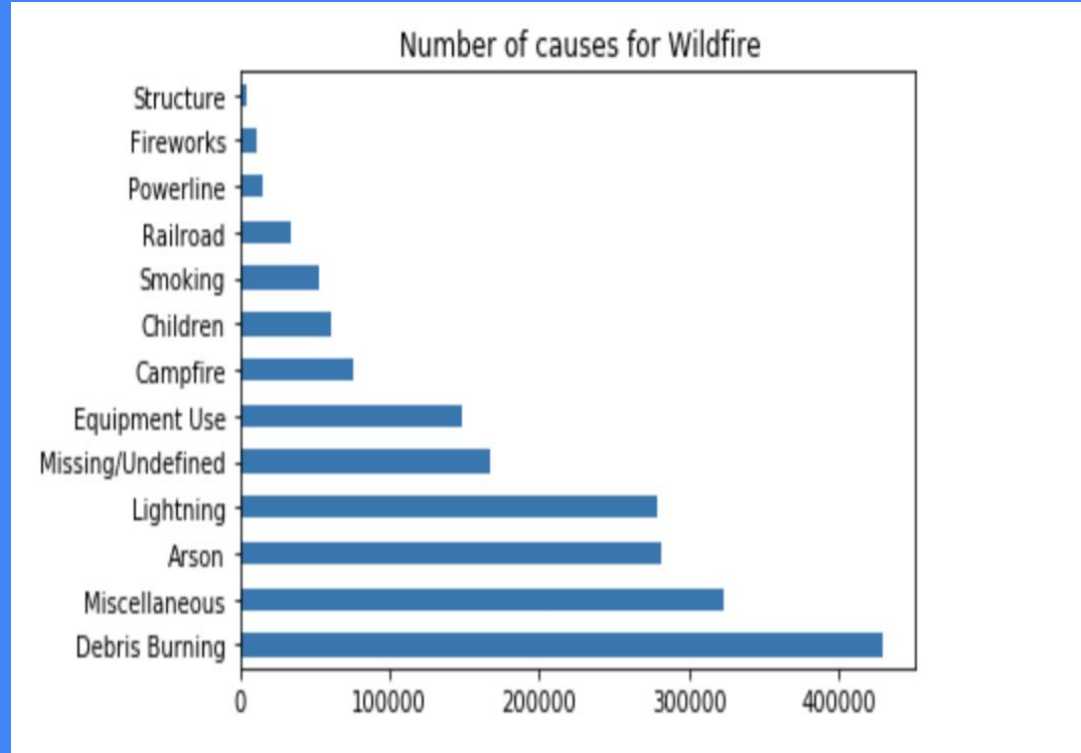
```
df_table = pd.read_sql_query("select * from sqlite_master where type = 'table' order by rootpage desc", conn)
df_table.head(5)
```

	type	name	tbl_name	rootpage	sql
0	table	NWCG_UnitIDActive_20170109	NWCG_UnitIDActive_20170109	776416	CREATE TABLE NWCG_UnitIDActive_20170109 (OBJEC...
1	table	idx_Fires_Shape_parent	idx_Fires_Shape_parent	677303	CREATE TABLE "idx_Fires_Shape_parent"(nodeno I...
2	table	idx_Fires_Shape_rowid	idx_Fires_Shape_rowid	677302	CREATE TABLE "idx_Fires_Shape_rowid"(rowid INT...
3	table	idx_Fires_Shape_node	idx_Fires_Shape_node	677301	CREATE TABLE "idx_Fires_Shape_node"(nodeno INT...
4	table	Fires	Fires	5673	CREATE TABLE Fires (OBJECTID integer primary k...

	FIRE_YEAR	NWCG_REPORTING_AGENCY	STAT_CAUSE_DESCR	FIRE_SIZE	LATITUDE	LONGITUDE	STATE	DISCOVERY_DATE
0	2005	FS	Miscellaneous	0.10	40.036944	-121.005833	CA	2453403.5
1	2004	FS	Lightning	0.25	38.933056	-120.404444	CA	2453137.5
2	2004	FS	Debris Burning	0.10	38.984167	-120.735556	CA	2453156.5
3	2004	FS	Lightning	0.10	38.559167	-119.913333	CA	2453184.5
4	2004	FS	Lightning	0.10	38.559167	-119.933056	CA	2453184.5

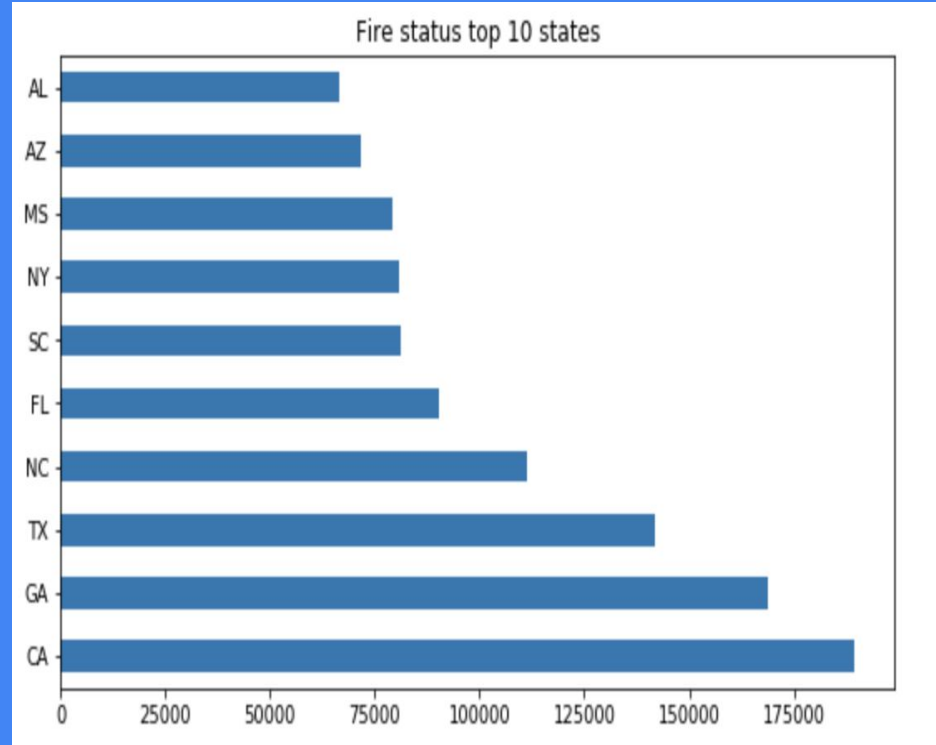
DATA EXPLORATION AND ANALYSIS

- 13 causes (classes) for wildfire.
- Divided four class
- Accident , Crime (Arson), natural(lightning), other
- Debris Burning
- Lightning and Arson are equal
- Structure less



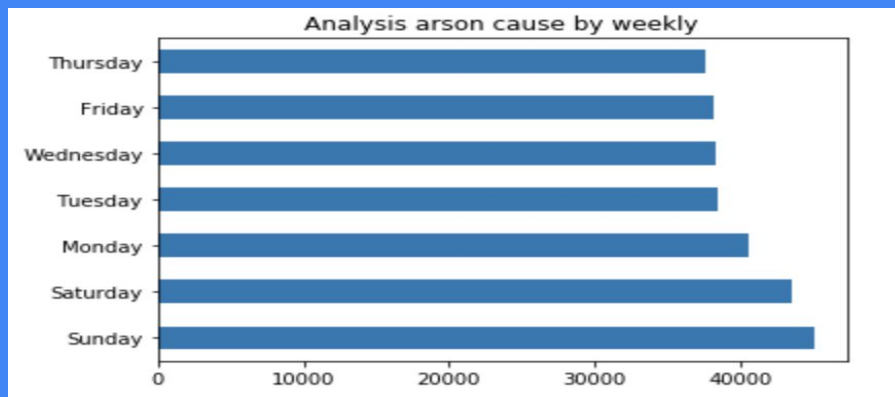
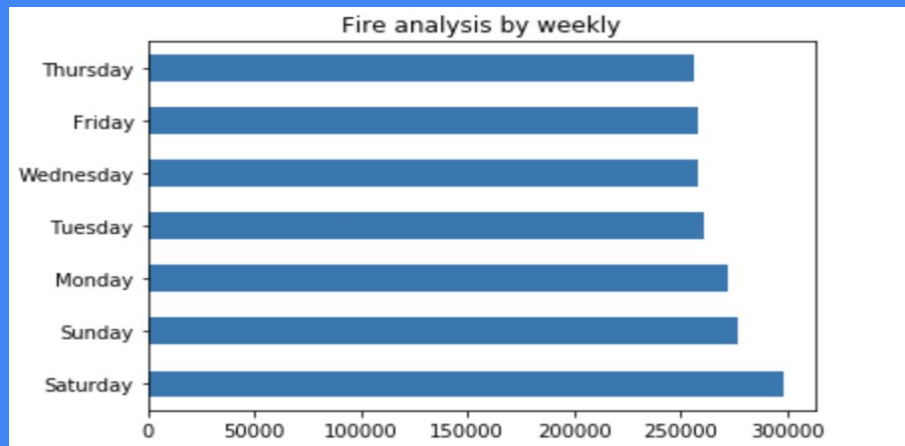
DATA EXPLORATION AND ANALYSIS...

- 10 ten states
- CA , GA ,TX are top three on Wildfire.



DATA EXPLORATION AND ANALYSIS...

- Saturday more wildfire
- Arson is main cause in Weekend



Model and Evaluation

- Logistic Regression
- KNearest neighbor algorithm
- Decision Tree
- Naive Bayes
- Random Forests

Model and Evaluation...

Modeling for four class (Natural , Accident ,Crime , Other)

Train Data Accuracy Score

```
LR Accuracy_score:0.402468 std:(0.005459)
KNN Accuracy_score:0.739594 std:(0.032133)
CART Accuracy_score:0.693515 std:(0.031033)
NB Accuracy_score:0.466783 std:(0.013459)
RN Accuracy_score:0.788153 std:(0.034717)
```

Test Data Accuracy Score

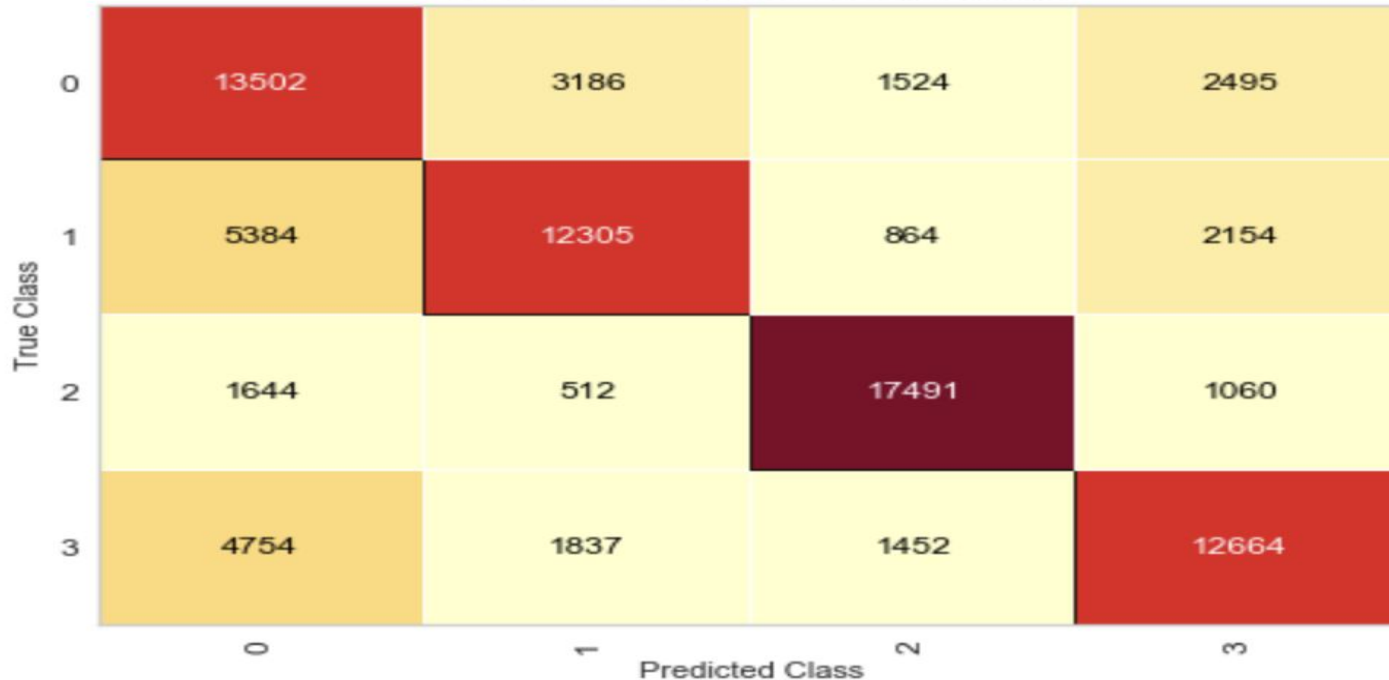
```
LR test data accuracy_score:0.404851
KNN test data accuracy_score:0.597117
CART test data accuracy_score:0.586637
NB test data accuracy_score:0.47426
RN test data accuracy_score:0.675315
```

Model and Evaluation...

Confusion Matrix for multi classification

RN test data accuracy_score:0.675641

RandomForestClassifier Confusion Matrix



Model and Evaluation ...

One class target Model evaluation - (Arson)

Train Data accuracy score , f1 score , Recall score and precision score

```
LR Accuracy_score:0.598932 std:(0.003940) ,f1_score: 0.623203 ,Recall : 0.663391 , Precision: 0.587622
KNN Accuracy_score:0.845204 std:(0.007241) ,f1_score: 0.858527 ,Recall : 0.939795 , Precision: 0.790268
CART Accuracy_score:0.815168 std:(0.013296) ,f1_score: 0.818227 ,Recall : 0.833356 , Precision: 0.803994
NB Accuracy_score:0.634828 std:(0.002618) ,f1_score: 0.672877 ,Recall : 0.751180 , Precision: 0.609373
RN Accuracy_score:0.884678 std:(0.014565) ,f1_score: 0.887335 ,Recall : 0.910385 , Precision: 0.865768
```

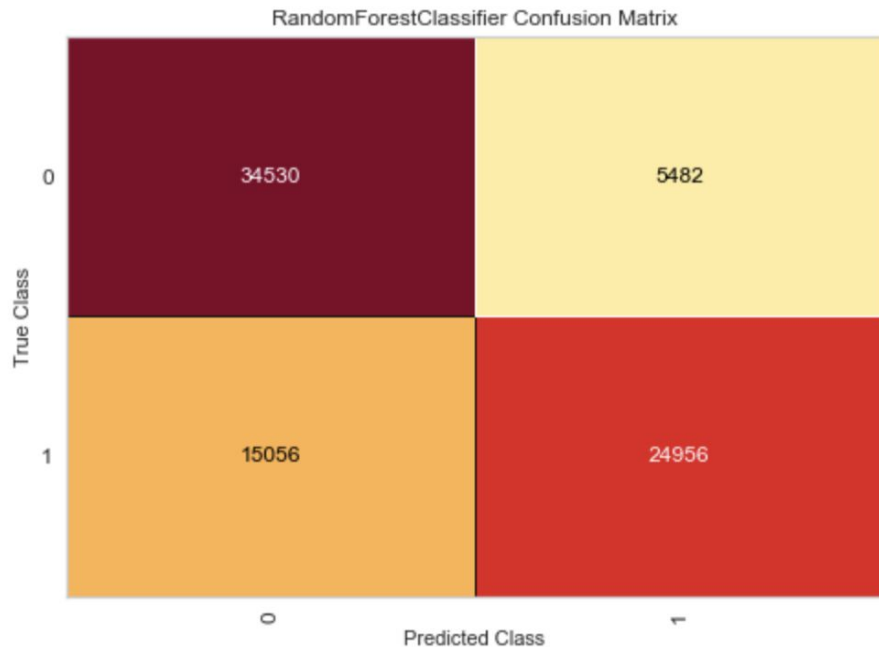
Test Data accuracy score , f1 score , Recall score and precision score

```
LR test data accuracy_score:0.60152 f1_score:0.62698 recall_score:0.669774 Precision score:0.589326
KNN test data accuracy_score:0.706051 f1_score:0.691473 recall_score:0.658802 Precision score:0.727554
CART test data accuracy_score:0.699053 f1_score:0.664362 recall_score:0.595696 Precision score:0.750922
NB test data accuracy_score:0.633785 f1_score:0.672676 recall_score:0.752599 Precision score:0.608098
RN test data accuracy_score:0.74214 f1_score:0.706668 recall_score:0.621214 Precision score:0.819384
```

Confusion Matrix

- Random forest Algorithm
- Accuracy score : 0.743 but Precision score : 0.819896

RN test data accuracy_score:0.743352 f1_score:0.708474 recall_score:0.623713 Precision score:0.819896



Conclusion

- We can predict arson cause in future wildfire .
- RandomForest algorithm
- Train accuracy score (84.47 %) and test accuracy score (74.22%)
- Test precision score (86.58%) and test data accuracy score (81.94%)

- Most Predictable Cause is **accident** among Natural , Accident, Crime, Other .
- Train Data Accuracy (**78.81%**)
- Test Data Accuracy (**67.53%**)

Thank You

Questions ?