Lower Back Pain Symptoms Dataset

Classification of backpain Normal or Abnormal for an IOT device.

Introduction

- Backpain is major concern for our society due to prolonged work hours. The requirement is to develop a model which is less compute intensive and provide good accuracy. The model is designed to deploy on a IOT device, hence the size of the model is critical for this project.
- Due to this requirement, we will be evaluating models which only saves weights the number of weights should be minimum.

Data Source

- The source of data is from a Kaggle repository. The url is https://www.kaggle.com/sammy123/lower-back-pain-symptoms-dataset

Features and label

X

- The features are
- pelvic_incidence
- pelvic tilt
- lumbar_lordosis_angle
- sacral_slope
- pelvic_radius
- degree_spondylolisthesis
- pelvic_slope
- direct_tiltthoracic_slope
- cervical_tilt
- sacrum_angle
- scoliosis_slope

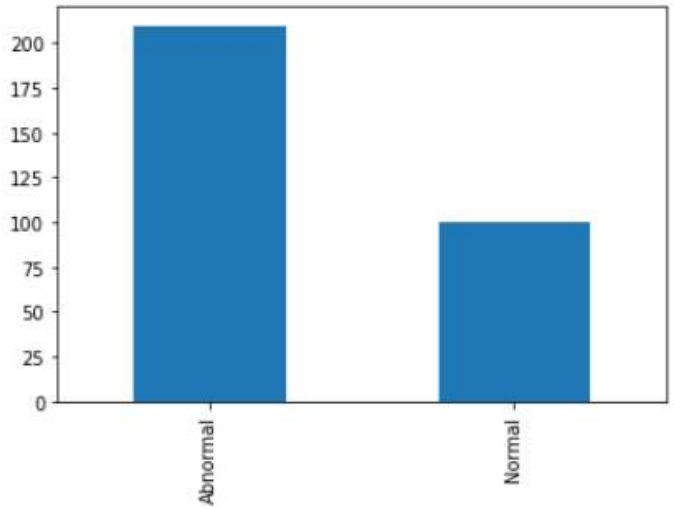
y

- Class_att

The target class values are imballanced

In [9]:
```
1 df["class_att"].value_counts().sort_index().plot.bar()
```

Out[9]: <AxesSubplot:>

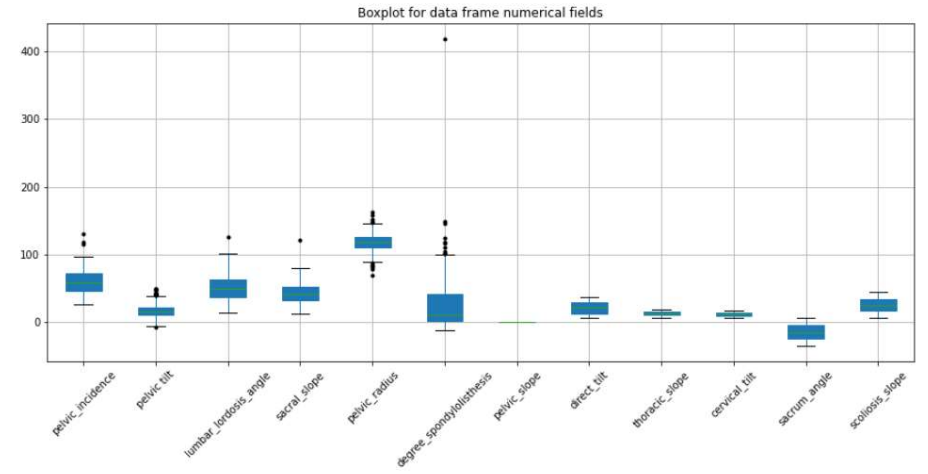


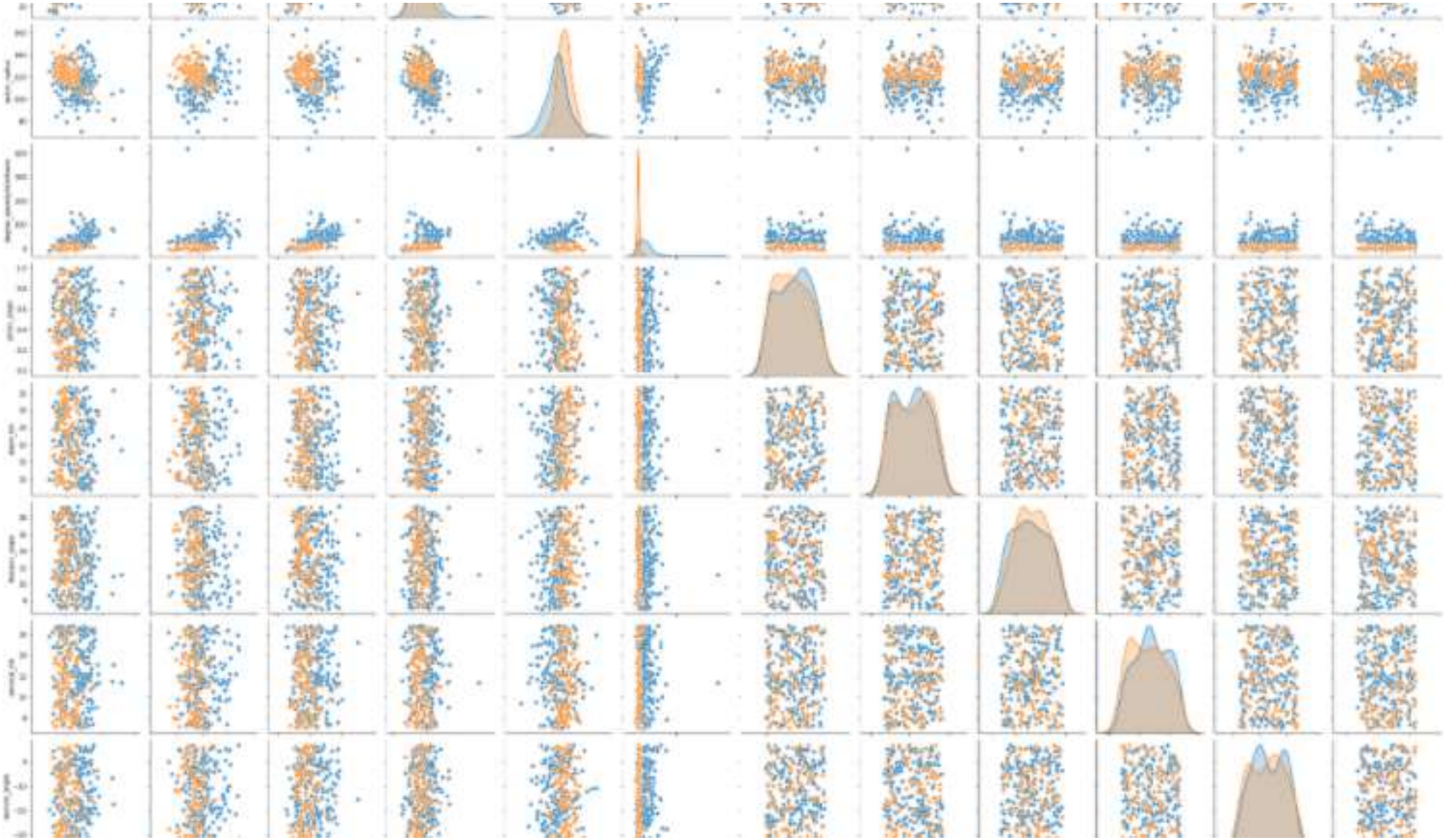
Box plot

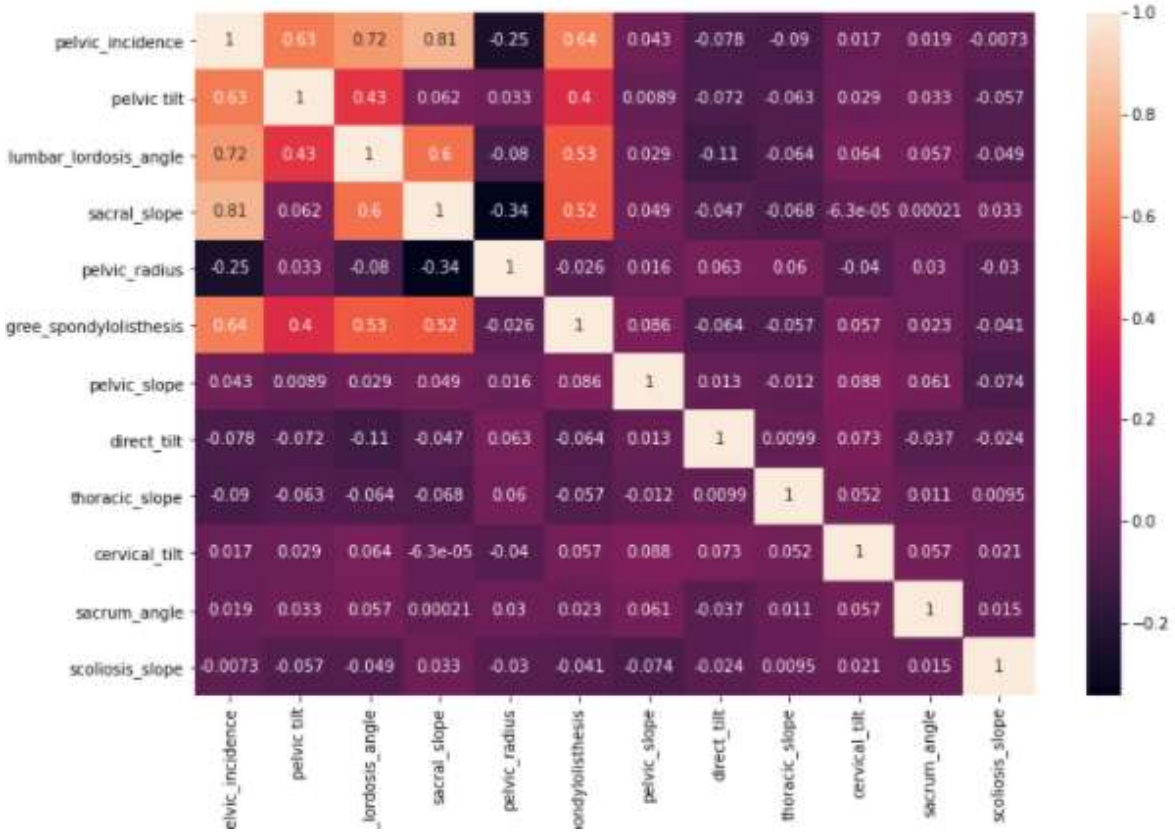
•

We can see few outliers in the dataset

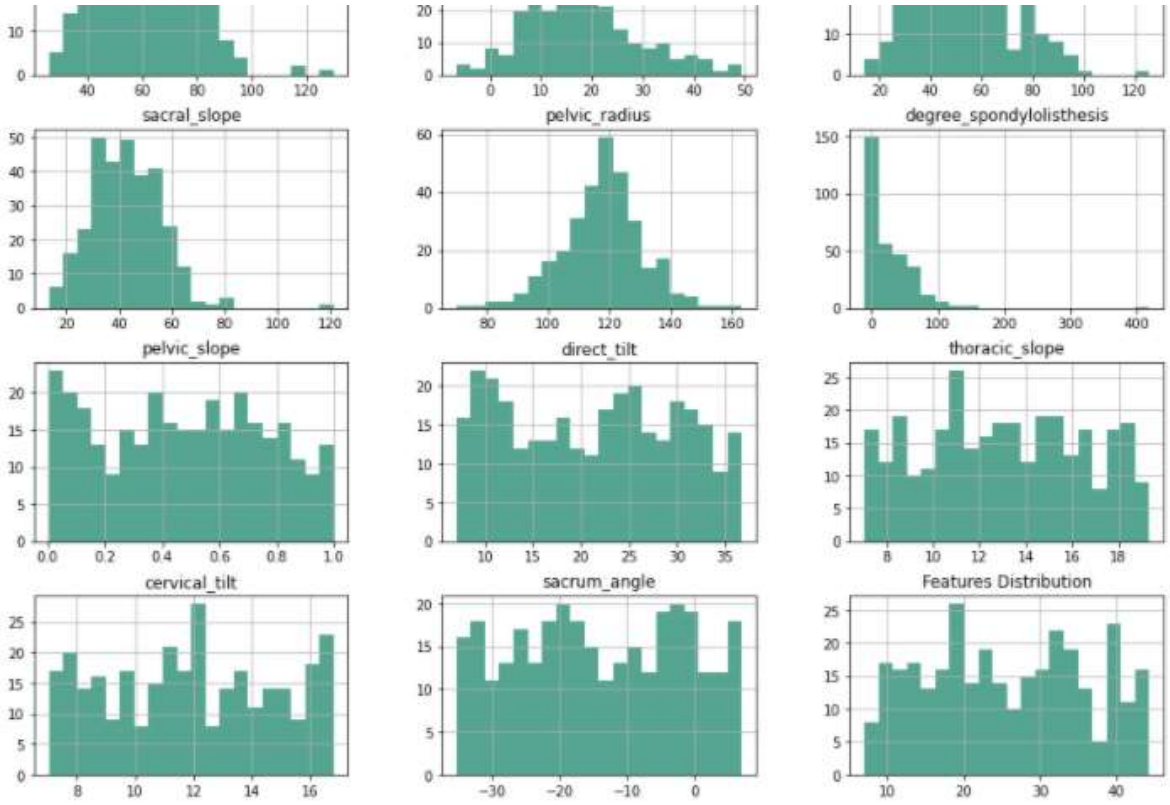
In [12]:
```
plt.subplots(figsize=(15,6))
df.boxplot(patch_artist=True, sym="k.")
plt.title("Boxplot for data frame numerical fields")
plt.xticks(rotation=45);
```







Correlation matrix



Feature distribution

SVC Classifier

In [21]:
```
from sklearn.pipeline import Pipeline
from sklearn.svm import SVC
```

In [22]:
```
pipesvc = Pipeline([('scaler', StandardScaler()), ('svc', SVC())]
pipesvc.fit(X_train,y_train)
pipesvc.score(X_test, y_test)
y_pred = pipesvc.predict(X_test)
```

In [23]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test, y_pred)
```

Out[23]:
```
array([[36,  6],
       [ 6, 14]], dtype=int64)
```

In [24]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

    Abnormal       0.86      0.86      0.86        42
      Normal       0.70      0.70      0.70        20

    accuracy                           0.81        62
   macro avg       0.78      0.78      0.78        62
weighted avg       0.81      0.81      0.81        62
```

SVC Classifier

DecisionTree classifier

In [25]:
```
from sklearn.tree import DecisionTreeClassifier
```

In [26]:
```
pipedtc = Pipeline([('scaler', StandardScaler()), ('svc', D
pipedtc.fit(X_train,y_train)
pipedtc.score(X_test, y_test)
y_pred = pipedtc.predict(X_test)
```

In [27]:
```
from sklearn.metrics import confusion_matrix
confusion_matrix(y_test, y_pred)
```

Out[27]:
```
array([[33,  9],
       [ 5, 15]], dtype=int64)
```

In [28]:
```
from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

    Abnormal       0.87      0.79      0.82        42
      Normal       0.62      0.75      0.68        20

    accuracy                           0.77        62
   macro avg       0.75      0.77      0.75        62
weighted avg       0.79      0.77      0.78        62
```

Decision Tree Classifier

Size on Disk

Size of SVC classifier on disk

```
]:  1 from joblib import dump, load
    2 dump(pipesvc, 'pipesvc.pkl')
    3 !dir pipesvc.pkl
```

```
 Volume in drive C is OS
 Volume Serial Number is AC8D-7FA8

 Directory of C:\Users\Ashok_Potti\Downloads\Personal\
ence Capstone

08/26/2020  11:18 PM            20,608 pipesvc.pkl
               1 File(s)         20,608 bytes
               0 Dir(s)  216,671,903,744 bytes free
```

of Decision Tree Classifier on di

```
np(pipedtc, 'pipedtc.pkl')
.r pipedtc.pkl
```

```
 in drive C is OS
 Serial Number is AC8D-7FA8

ory of C:\Users\Ashok_Potti\Downloads\Personal\Cours
pstone

020  11:18 PM             6,991 pipedtc.pkl
       1 File(s)          6,991 bytes
       0 Dir(s)  216,671,903,744 bytes free
```

Conclusion

- **The accuracy of SVC classifier is 4% more than Decision Tree Classifier.**
- **But the size of SVC is 20k bytes compared to 7k by Decision tree classifier.**
- **Hence for premium device SVC model will be deployed and for mid-range device Decision Tree classifier will be deployed.**

- Thank you