Motivation

- Soccer players transfer clubs very often today and their market values can vary due to a number of different factors.
- Use MLlib based analysis to identify the most important factor that correlates with a player’s market value.
- Build a machine learning model that can predict a player’s market value based on player’s statistics.

Pearson correlation coefficient of the regressive factors

```
print("Pearson's r(fpl_value,market_value) = {}".format(df.corr("fpl_value", "market_value")))
print("Pearson's r(fpl_points,market_value) = {}".format(df.corr("fpl_points", "market_value")))
print("Pearson's r(page_views,market_value) = {}".format(df.corr("page_views", "market_value")))
```

```
Pearson's r(fpl_value,market_value) = 0.7869349160777618
Pearson's r(fpl_points,market_value) = 0.5374665750584899
Pearson's r(page_views,market_value) = 0.6765409830507209
```

Sample of the prediction made by machine learning model

```
predictionsAndLabelsDF = lrModel.transform(testDF)

print(predictionsAndLabelsDF.orderBy(predictionsAndLabelsDF.label.desc()).take(5))
```

```
[Row(label=50.0, features=DenseVector([11.5]), prediction=54.56564999133577), Row(label=45.0, features=DenseVector
([9.0]), prediction=36.63693716164054), Row(label=35.0, features=DenseVector([9.0]), prediction=36.63693716164054), R
ow(label=35.0, features=DenseVector([7.0]), prediction=22.293966897884353), Row(label=30.0, features=DenseVector([8.
0]), prediction=29.465452029762446)]
```

Evaluation of the model

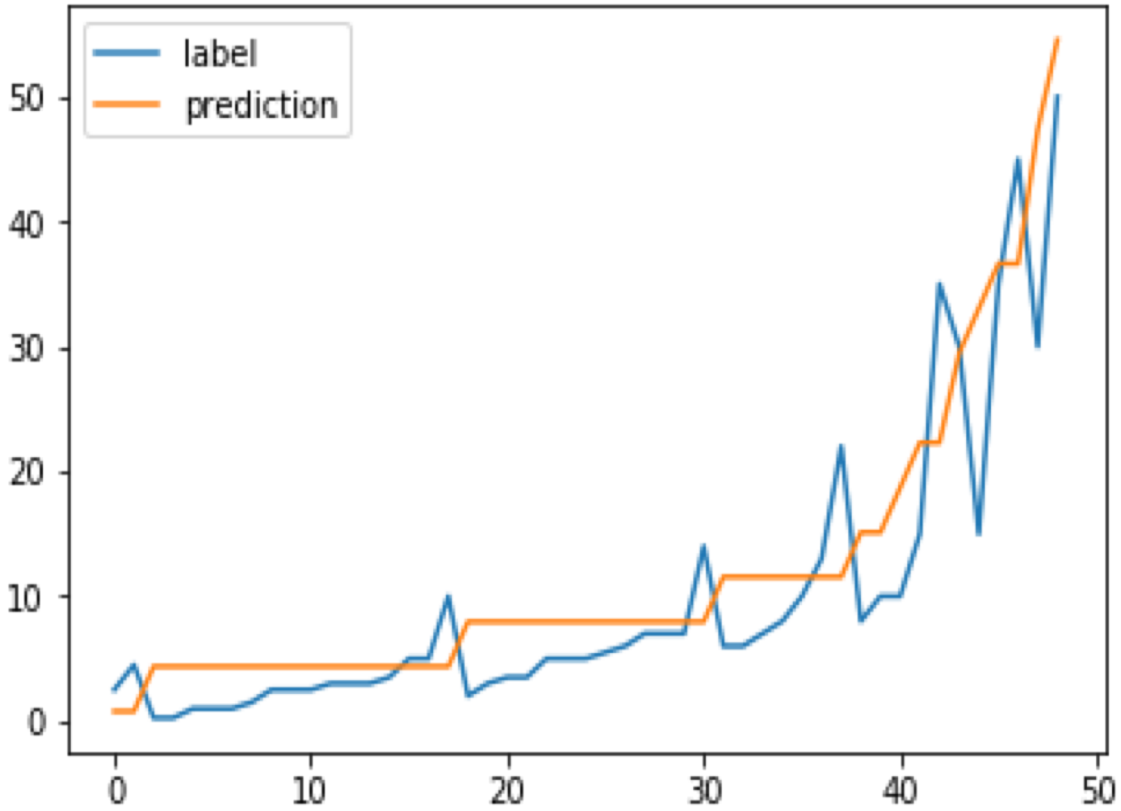
```
eval.setMetricName("rmse").evaluate(predictionsAndLabelsDF)
```

```
5.74095422393865
```

```
eval.setMetricName("r2").evaluate(predictionsAndLabelsDF)
```

```
0.7506525733598287
```

Data Visualization



Market_value vs. predicted result produced by model using test dataset