With modification 1, 2 Results

```
[root@f814c003cacc:/tmp/ml# python3 ml_pyspark.py                              ]
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/05/23 04:17:36 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
Area Under ROC Curve: 0.8948
Selected Maximum Tree Depth: 6
root@f814c003cacc:/tmp/ml#
```

Area Under ROC Curve: 0.8948
Selected Maximum Tree Depth: 6

W/O modification 1, 2 Results

```
[root@3304de382529:/tmp/ml# python3 w_o_changes_ml_pyspark.py
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/05/23 04:47:28 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... us
ing builtin-java classes where applicable
Area Under ROC Curve: 0.8708
Selected Maximum Tree Depth: 4
root@3304de382529:/tmp/ml#
```

Area Under ROC Curve: 0.8708
Selected Maximum Tree Depth: 4

Comparison Analysis

The model with modifications shows an AUC ROC of 0.8948, which is significantly higher than the 0.8708 AUC ROC of the model without modifications. This indicates an improvement in the model's ability to correctly classify the positive and negative classes. The improvements from the modifications suggest that the changes made have enhanced the model's predictive power. This makes sense as squaring the features could help the model capture non-linear relationships between the predictors and the target variable. There's also a notable difference in the selected maximum tree depths between the two models. The model with modifications selected a deeper tree (depth of 6) compared to the model without modifications (depth of 4). This could imply that the modifications allow the model to leverage deeper, more complex tree structures effectively, possibly because the modifications help in handling overfitting or providing richer information for splits at deeper levels of the trees. In conclusion, the modifications have not only improved the ROC AUC but also influenced the complexity of the model that achieves the best performance. This result suggests that the modifications likely introduced new information or interactions that the model could exploit to make more nuanced decisions at deeper levels of the tree.