Injury Prediction for Competitive Runners

Rawan Mansour

Abdulwahab Al-Subit

Table of Contents

Introduction

Staying injury free is a major factor for success in sports. Our purpose was to use machine learning for the prediction of injuries in runners, based on detailed training logs.

Dataset

Injury Prediction In Competitive RunnersDataset

provided by Kaggle website

- Dataset was obtained from kaggle in form (.csv)
- Has 13 columns and 42766 rows.
- Include a binary column indicating whether

  this training setup resulted in an injury (1) or not (0).
- The target I want to predict is “injury”.

Cleaning data

01

Missing

Filling the missing value

02

Unnecessary

Drop the unnecessary features

03

Duplicate

Check for duplicates data

04

Type

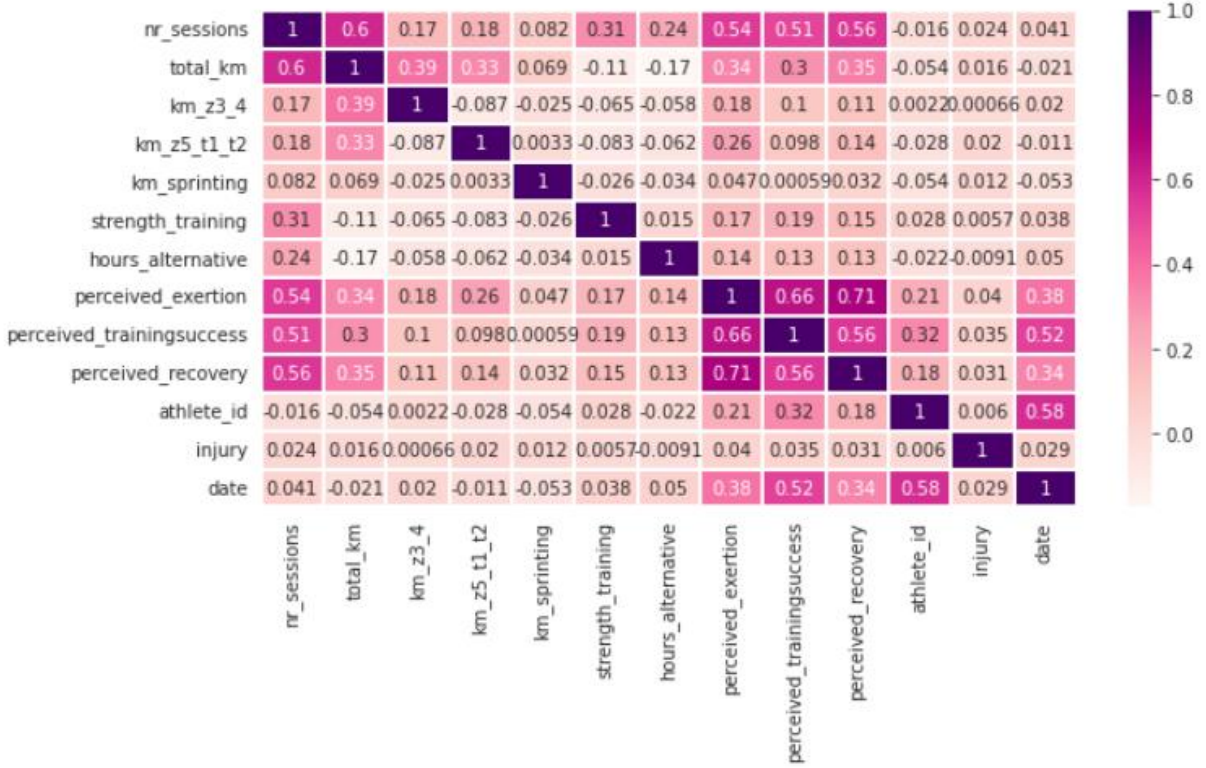
Checking features type

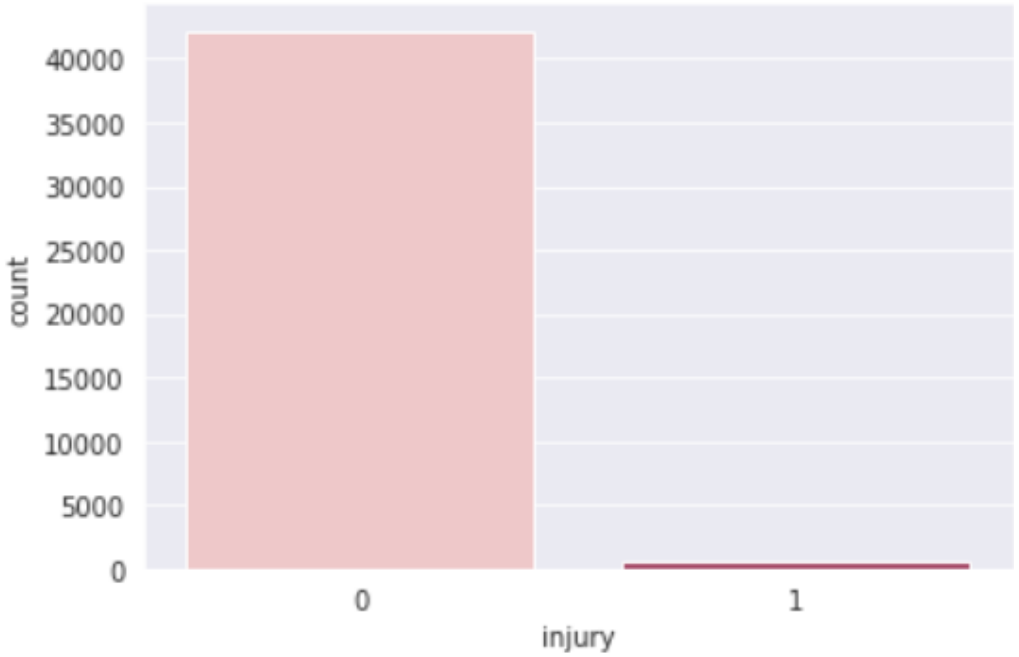
Visualization

Exploring Data Analysis

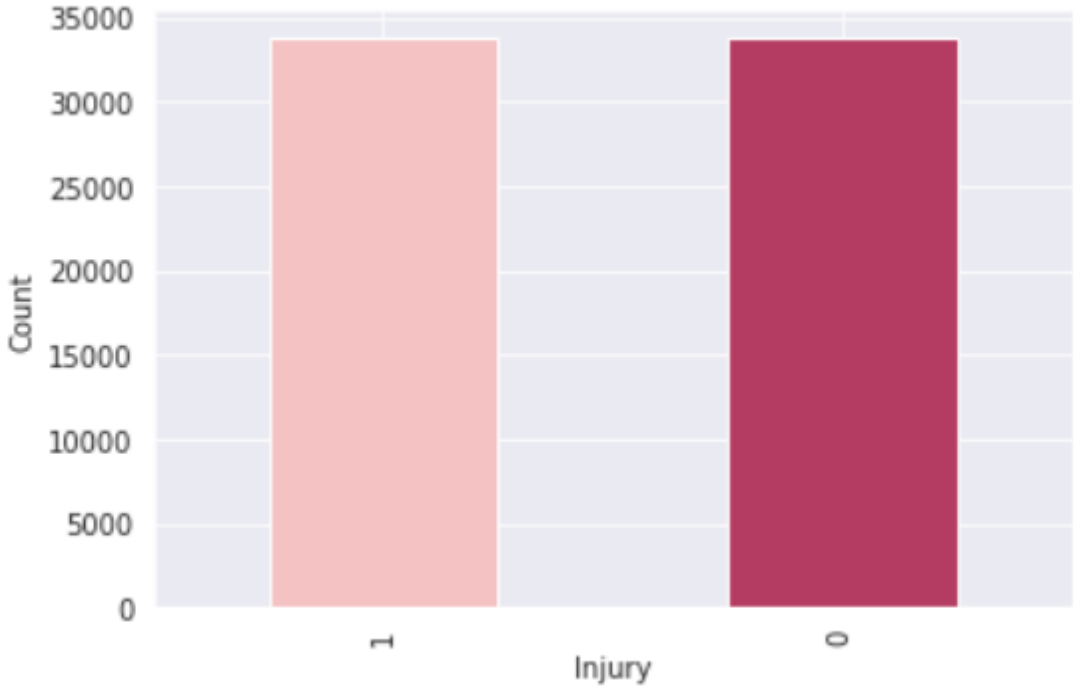
(EDA)

The correlation heatmap was used to find the correlation between factors





This plot shows there is unbalance between in the dataset so, in preprocessing part I will balance between them using SMOTE.



After balancing data.

The Model Used

- Logistic Regression.
- Xgboost.
- KNN.

Comparison Models

	xgboost	Logistic Regression	KNN
Accuracy	86%	62%	98%
Recall	0.63	0.87	0.96
f1-score	0.62	0.86	0.98

Conclusion

KNN

Is the best model to predict the possibility of Injury for Competitive Runners.

Tools Used

- For Data Processing

  - Pandas.

  - NumPy.

- For Building The Model

  - Scikit-learn library.

  - Imblearn.

  - XGBoost

- For Visualization

  - Matplotlib.

  - Seaborn.

Thanks!

Do you have any questions?