

Article

An Exploratory Study of ML Techniques in Football Match's Result Prediction

Vivek Jain¹, Prathu Bhati¹, Rajbali Singh¹, Raj Maurya¹, Rachit Joshi¹

¹Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh, India.

I N F O

Corresponding Author:

Prathu Bhati, Department of Computer Science and Engineering, IMS Engineering College, Ghaziabad, Uttar Pradesh, India.

E-mail Id:

prathubhati184@gmail.com

Orcid Id:

<https://orcid.org/0000-0003-0698-3370>

How to cite this article:

Jain V, Bhati P, Singh R et al. An Exploratory Study of ML Techniques in Football Match's Result Prediction. *J Adv Res Embed Sys* 2022; 9(3&4): 1-4.

Date of Submission: 2022-12-22

Date of Acceptance: 2023-01-10

A B S T R A C T

Machine learning is a subset of artificial intelligence (AI) that allows computers to learn and improvise on their own without having to be explicitly programmed. Machine learning deals with the creation of computer programs that can access data and learn on their own. Sports prediction is one of the rapidly-growing fields in good predictive accuracy since it involves a large sum of money in betting. The capability to apply algorithms and use that knowledge to try to forecast the outcome of future games based on this data is a particularly important aspect of machine learning in football. Sports match results can be difficult to forecast, with unexpected outcomes frequently occurring. Football is a good example since matches have a set length (as opposed to racket sports like tennis, where the game is played until one player wins). In this study, Machine Learning techniques are used to predict the winning team in the English Premier League (EPL). The goal is to predict a football match's Full-Time Result (FTR) accurately, which determines the winning team. For training the data, we use algorithms like Support Vector Machines, XG Boost, Logistic Regression, the one with the highest and best accuracy is used to forecast the winning team. The data for previous seasons is obtained from.⁶

Keywords: Football, Soccer Analytics, Prediction, Machine Learning, Support Vector Machine (SVM), XG Boost.

Introduction

Football is the most popular sport globally and is played by 250 million players in over 200 countries. Analytics has always been there in the field of sports even if we don't acknowledge it.

To be more precise, analytics in the field of football is the method of creating meaningful information and decisions that can be acted upon using soccer-related data. The data includes anything ranging from how many goals a team has scored to multiple factors like, distance covered by a player during the course of the match, or a number of passes played and how many out of those were accurate

along with how many out of those created a chance for their team to score and so on.

In every soccer, league groups are formed and the teams play 2 matches with each alternative team in their league - one at their home structure and the other at the opponent's home stadium. Every such match has 3 doable outcomes the home side wins, the match ends in a draw, or the visiting team wins.

Given such a format, it's natural that there are many online fantasy leagues, betting agencies, others who attempt to predict the end result of every match. during this project, an endeavor has been created to seek out the factors that have an effect on the outcome of a match and conjointly

predict the results of any fixture by utilizing these factors.

The most important reason in the back of this venture is giving a correct dataset for football matches and predicting the winners in upcoming games and hence yielding efficient results. In this paper, we suggest a version of football prediction primarily based totally on FTR that is Full-Time Result (Our Class label) i.e. Home, Away or Draw.

Background

Outcomes from sports matches is difficult to predict, with surprises often doping up. Football specifically is a noteworthy example as matches have fixed length (as opposition racket sports like tennis, where the sport is played until a player wins).

However, because of the low-scoring nature of games (less than 3 goals per game on the average within the english premier league within the past 15 years) there's a random element linked to the quantity of goals scored during a match. There is a necessity to seek out if the appliance of machine learning can bring better and more insightful lead to soccer analytics. This makes match results an imperfect measure of a team's performance and thus an incomplete metric on which to predict future results.

Literature Survey

Thamaraimanalan et al.¹ utilized the object detection within the football games depends upon the basic image processing techniques. They carried out training process using Convolutional Neural Network (CNN) classifier with the accuracy rate of 87.63% with the reduction in the sensitivity and the specificity range of 72.5 and 86.2%.

Yang² in their work proposed a Linear Support Vector Classifier (LSVC) to predict the outcome of the match based on the performance of the players. Model was validated with the results of the statistics of the AUC, F1 and prediction accuracy of the model were 0.8597, 0.6973 and 0.7965 respectively on the verification data.

Martinovic J, Snásel V, Ochodkova E, Nolta L, Wu J, Abraham A³ in their work studied robot soccer game, as a part of standard applications of distributed system control in real time to predict strategy and perform game analysis. Uhrín Matej, Šourek Gustav, Hubáček Ondřej, Železný Filip⁴ investigated the two most prominent streams of betting investment strategies based on the views of the Modern Portfolio Theory and the Kelly criterion, together with a number of their popular modifications aimed at additional risk management in practice, where their original underlying mathematical assumptions do not hold.

Woo-Joo Lee, Hyo-Jin Jhang, Seung Hoe Choi⁵ performed a study that aims to find variables that affect the winning rate of the football team before a match. Qualitative variables such as venue, match importance, performance,

atmosphere of both teams are suggested to predict the outcome Using Regression analysis.

Edward Wheatcroft⁶ studied use of observed and predicted match statistics as inputs to forecasts for the outcomes of football matches. It is shown that, were it possible to know the match statistics in advance, highly informative forecasts of the match outcome could be made.

Rahul Baboota, Harleen Kaur⁷ used feature engineering and exploratory data analysis to create a feature set for determining the most important factors for predicting the results of a football match, consequently create a highly accurate predictive system using machine learning. Using gradient boosting they achieved a performance of 0.2156 on the Ranked Probability Score (RPS) metric for game weeks 6 to 38 for the English Premier League aggregated over two seasons (2014-2015 and 2015-2016), whereas the betting organizations that we consider (Bet 365 and Pinnacle Sports) obtained an RPS value of 0.2012 for the same period.

Proposed Model

We present a model to predict the outcome of football matches in the English Premier League. We prepare the dataset of past seasons on various machine learning classifiers. Comparisons amongst the algorithms would be made and the one that turns out to be the most precise i.e. having the sounder forecast accuracy will be considered. Then, optimization can be produced on that classifier to further enhance the model's precision in making forecasts. The tag that would be considered would be Home Win (H), Away Win (A), Draw (D).

Dataset Description

The forecast is done based on data from past games for recent seasons. We have obtained the data set from⁸ that has an enormous quantity of data right from the old games to the ones that are being played. There are about 65 features per season like the Home team, Away team, scores, venue to be named a few. After including filtered these features we get about 8-10 features that are going to foresee the outcomes. The dataset size is 6080.

#	Date	HomeTeam	AwayTeam	FTR	FTAG	FTR	HTAG	ATAG	HTP	ATP	HM1	HM2	HM3	HM4	HM5	HM6	HM7	HM8	HM9	HM10	HM11	HM12	HM13	HM14	HM15	HM16	HM17	HM18	HM19	HM20	HM21	HM22	HM23	HM24	HM25	HM26	HM27	HM28	HM29	HM30	HM31	HM32	HM33	HM34	HM35	HM36	HM37	HM38	HM39	HM40	HM41	HM42	HM43	HM44	HM45	HM46	HM47	HM48	HM49	HM50	HM51	HM52	HM53	HM54	HM55	HM56	HM57	HM58	HM59	HM60	HM61	HM62	HM63	HM64	HM65	HM66	HM67	HM68	HM69	HM70	HM71	HM72	HM73	HM74	HM75	HM76	HM77	HM78	HM79	HM80	HM81	HM82	HM83	HM84	HM85	HM86	HM87	HM88	HM89	HM90	HM91	HM92	HM93	HM94	HM95	HM96	HM97	HM98	HM99	HM100	HM101	HM102	HM103	HM104	HM105	HM106	HM107	HM108	HM109	HM110	HM111	HM112	HM113	HM114	HM115	HM116	HM117	HM118	HM119	HM120	HM121	HM122	HM123	HM124	HM125	HM126	HM127	HM128	HM129	HM130	HM131	HM132	HM133	HM134	HM135	HM136	HM137	HM138	HM139	HM140	HM141	HM142	HM143	HM144	HM145	HM146	HM147	HM148	HM149	HM150	HM151	HM152	HM153	HM154	HM155	HM156	HM157	HM158	HM159	HM160	HM161	HM162	HM163	HM164	HM165	HM166	HM167	HM168	HM169	HM170	HM171	HM172	HM173	HM174	HM175	HM176	HM177	HM178	HM179	HM180	HM181	HM182	HM183	HM184	HM185	HM186	HM187	HM188	HM189	HM190	HM191	HM192	HM193	HM194	HM195	HM196	HM197	HM198	HM199	HM200	HM201	HM202	HM203	HM204	HM205	HM206	HM207	HM208	HM209	HM210	HM211	HM212	HM213	HM214	HM215	HM216	HM217	HM218	HM219	HM220	HM221	HM222	HM223	HM224	HM225	HM226	HM227	HM228	HM229	HM230	HM231	HM232	HM233	HM234	HM235	HM236	HM237	HM238	HM239	HM240	HM241	HM242	HM243	HM244	HM245	HM246	HM247	HM248	HM249	HM250	HM251	HM252	HM253	HM254	HM255	HM256	HM257	HM258	HM259	HM260	HM261	HM262	HM263	HM264	HM265	HM266	HM267	HM268	HM269	HM270	HM271	HM272	HM273	HM274	HM275	HM276	HM277	HM278	HM279	HM280	HM281	HM282	HM283	HM284	HM285	HM286	HM287	HM288	HM289	HM290	HM291	HM292	HM293	HM294	HM295	HM296	HM297	HM298	HM299	HM300	HM301	HM302	HM303	HM304	HM305	HM306	HM307	HM308	HM309	HM310	HM311	HM312	HM313	HM314	HM315	HM316	HM317	HM318	HM319	HM320	HM321	HM322	HM323	HM324	HM325	HM326	HM327	HM328	HM329	HM330	HM331	HM332	HM333	HM334	HM335	HM336	HM337	HM338	HM339	HM340	HM341	HM342	HM343	HM344	HM345	HM346	HM347	HM348	HM349	HM350	HM351	HM352	HM353	HM354	HM355	HM356	HM357	HM358	HM359	HM360	HM361	HM362	HM363	HM364	HM365	HM366	HM367	HM368	HM369	HM370	HM371	HM372	HM373	HM374	HM375	HM376	HM377	HM378	HM379	HM380	HM381	HM382	HM383	HM384	HM385	HM386	HM387	HM388	HM389	HM390	HM391	HM392	HM393	HM394	HM395	HM396	HM397	HM398	HM399	HM400	HM401	HM402	HM403	HM404	HM405	HM406	HM407	HM408	HM409	HM410	HM411	HM412	HM413	HM414	HM415	HM416	HM417	HM418	HM419	HM420	HM421	HM422	HM423	HM424	HM425	HM426	HM427	HM428	HM429	HM430	HM431	HM432	HM433	HM434	HM435	HM436	HM437	HM438	HM439	HM440	HM441	HM442	HM443	HM444	HM445	HM446	HM447	HM448	HM449	HM450	HM451	HM452	HM453	HM454	HM455	HM456	HM457	HM458	HM459	HM460	HM461	HM462	HM463	HM464	HM465	HM466	HM467	HM468	HM469	HM470	HM471	HM472	HM473	HM474	HM475	HM476	HM477	HM478	HM479	HM480	HM481	HM482	HM483	HM484	HM485	HM486	HM487	HM488	HM489	HM490	HM491	HM492	HM493	HM494	HM495	HM496	HM497	HM498	HM499	HM500	HM501	HM502	HM503	HM504	HM505	HM506	HM507	HM508	HM509	HM510	HM511	HM512	HM513	HM514	HM515	HM516	HM517	HM518	HM519	HM520	HM521	HM522	HM523	HM524	HM525	HM526	HM527	HM528	HM529	HM530	HM531	HM532	HM533	HM534	HM535	HM536	HM537	HM538	HM539	HM540	HM541	HM542	HM543	HM544	HM545	HM546	HM547	HM548	HM549	HM550	HM551	HM552	HM553	HM554	HM555	HM556	HM557	HM558	HM559	HM560	HM561	HM562	HM563	HM564	HM565	HM566	HM567	HM568	HM569	HM570	HM571	HM572	HM573	HM574	HM575	HM576	HM577	HM578	HM579	HM580	HM581	HM582	HM583	HM584	HM585	HM586	HM587	HM588	HM589	HM590	HM591	HM592	HM593	HM594	HM595	HM596	HM597	HM598	HM599	HM600	HM601	HM602	HM603	HM604	HM605	HM606	HM607	HM608	HM609	HM610	HM611	HM612	HM613	HM614	HM615	HM616	HM617	HM618	HM619	HM620	HM621	HM622	HM623	HM624	HM625	HM626	HM627	HM628	HM629	HM630	HM631	HM632	HM633	HM634	HM635	HM636	HM637	HM638	HM639	HM640	HM641	HM642	HM643	HM644	HM645	HM646	HM647	HM648	HM649	HM650	HM651	HM652	HM653	HM654	HM655	HM656	HM657	HM658	HM659	HM660	HM661	HM662	HM663	HM664	HM665	HM666	HM667	HM668	HM669	HM670	HM671	HM672	HM673	HM674	HM675	HM676	HM677	HM678	HM679	HM680	HM681	HM682	HM683	HM684	HM685	HM686	HM687	HM688	HM689	HM690	HM691	HM692	HM693	HM694	HM695	HM696	HM697	HM698	HM699	HM700	HM701	HM702	HM703	HM704	HM705	HM706	HM707	HM708	HM709	HM710	HM711	HM712	HM713	HM714	HM715	HM716	HM717	HM718	HM719	HM720	HM721	HM722	HM723	HM724	HM725	HM726	HM727	HM728	HM729	HM730	HM731	HM732	HM733	HM734	HM735	HM736	HM737	HM738	HM739	HM740	HM741	HM742	HM743	HM744	HM745	HM746	HM747	HM748	HM749	HM750	HM751	HM752	HM753	HM754	HM755	HM756	HM757	HM758	HM759	HM760	HM761	HM762	HM763	HM764	HM765	HM766	HM767	HM768	HM769	HM770	HM771	HM772	HM773	HM774	HM775	HM776	HM777	HM778	HM779	HM780	HM781	HM782	HM783	HM784	HM785	HM786	HM787	HM788	HM789	HM790	HM791	HM792	HM793	HM794	HM795	HM796	HM797	HM798	HM799	HM800	HM801	HM802	HM803	HM804	HM805	HM806	HM807	HM808	HM809	HM810	HM811	HM812	HM813	HM814	HM815	HM816	HM817	HM818	HM819	HM820	HM821	HM822	HM823	HM824	HM825	HM826	HM827	HM828	HM829	HM830	HM831	HM832	HM833	HM834	HM835	HM836	HM837	HM838	HM839	HM840	HM841	HM842	HM843	HM844	HM845	HM846	HM847	HM848	HM849	HM850	HM851	HM852	HM853	HM854	HM855	HM856	HM857	HM858	HM859	HM860	HM861	HM862	HM863	HM864	HM865	HM866	HM867	HM868	HM869	HM870	HM871	HM872	HM873	HM874	HM875	HM876	HM877	HM878	HM879	HM880	HM881	HM882	HM883	HM884	HM885	HM886	HM887	HM888	HM889	HM890	HM891	HM892	HM893	HM894	HM895	HM896	HM897	HM898	HM899	HM900	HM901	HM902	HM903	HM904	HM905	HM906	HM907	HM908	HM909	HM910	HM911	HM912	HM913	HM914	HM915	HM916	HM917	HM918	HM919	HM920	HM921	HM922	HM923	HM924	HM925	HM926	HM927	HM928	HM929	HM930	HM931	HM932	HM933	HM934	HM935	HM936	HM937	HM938	HM939	HM940	HM941	HM942	HM943	HM944	HM945	HM946	HM947	HM948	HM949	HM950	HM951	HM952	HM953	HM954	HM955	HM956	HM957	HM958	HM959	HM960	HM961	HM962	HM963	HM964	HM965	HM966	HM967	HM968	HM969	HM970	HM971	HM972	HM973	HM974	HM975	HM976	HM977	HM978	HM979	HM980	HM981	HM982	HM983	HM984	HM985	HM986	HM987	HM988	HM989	HM990	HM991	HM992	HM993	HM994	HM995	HM996	HM997	HM998	HM999	HM1000	HM1001	HM1002	HM1003	HM1004	HM1005	HM1006	HM1007	HM1008	HM1009	HM1010	HM1011	HM1012	HM1013	HM1014	HM1015	HM1016	HM1017	HM1018	HM1019	HM1020	HM1021	HM1022	HM1023	HM1024	HM1025	HM1026	HM1027	HM1028	HM1029	HM1030	HM1031	HM1032	HM1033	HM1034	HM1035	HM1036	HM1037	HM1038	HM1039	HM1040	HM1041	HM1042	HM1043	HM1044	HM1045	HM1046	HM1047	HM1048	HM1049	HM1050	HM1051	HM1052	HM1053	HM1054	HM1055	HM1056	HM1057	HM1058	HM1059	HM1060	HM1061	HM1062	HM1063	HM1064	HM1065	HM1066	HM1067	HM1068	HM1069	HM1070	HM1071	HM1072	HM1073	HM1074	HM1075	HM1076	HM1077	HM1078	HM1079	HM1080	HM1081	HM1082	HM1083	HM1084	HM1085	HM1086	HM1087	HM1088	HM1089	HM1090	HM1091	HM1092	HM1093	HM1094	HM1095	HM1096	HM1097	HM1098	HM1099	HM1100	HM1101	HM1102	HM1103	HM1104	HM1105	HM1106	HM1107	HM1108	HM1109	HM1110	HM1111	HM1112	HM1113	HM1114	HM1115	HM1116	HM1117	HM1118	HM1119	HM1120	HM1121	HM1122	HM1123	HM1124	HM1125	HM1126	HM1127	HM1128	HM1129	HM1130	HM1131	HM1132	HM1133	HM1134	HM1135	HM1136	HM1137	HM1138	HM1139	HM1140	HM1141	HM1142	HM1143	HM1144	HM1145	HM1146	HM1147	HM1148	HM1149	HM1150	HM1151	HM1152	HM1153	HM1154	HM1155	HM1156	HM1157	HM1158	HM1159	HM1160	HM1161	HM1162	HM1163	HM1164	HM1165	HM1166	HM1167	HM1168	HM1169	HM1170	HM1171	HM1172	HM1173	HM1174	HM1175	HM1176	HM1177	HM1178	HM1179	HM1180	HM1181	HM1182	HM1183	HM1184	HM1185	HM1186	HM1187	HM1188	HM1189	HM1190	HM1191	HM1192	HM1193	HM1194	HM1195	HM1196	HM1197	HM1198	HM1199	HM1200	HM1201	HM1202	HM1203	HM1204	HM1205	HM1206	HM1207	HM1208	HM1209	HM1210	HM1211	HM1212	HM1213	HM1214	HM1215	HM1216	HM1217	HM1218	HM1219	HM1220	HM1221	HM1222	HM1223	HM1224	HM1225	HM1226	HM1227	HM1228	HM1229	HM1230	HM1231	HM1232	HM1233	HM1234	HM1235	HM1236	HM1237	HM1238	HM1239	HM1240	HM1241	HM1242	HM1243	HM1244	HM1245	HM1246	HM1247	HM1248	HM1249	HM1250	HM1251	HM1252	HM1253	HM1254	HM1255	HM1256	HM1257	HM1258	HM1259	HM1260	HM1261	HM1262	HM1263	HM1264	HM1265	HM1266	HM1267	HM1268	HM1269	HM1270	HM1271	HM1272	HM1273	HM1274	HM1275	HM1276	HM1277	HM1278	HM1279	HM1280	HM1281	HM1282	HM1283	HM1284	HM1285	HM1286	HM1287	HM1288	HM1289	HM1290	HM1291	HM1292	HM1293	HM1294	HM1295	HM1296	HM1297	HM1298	HM1299	HM1300	HM1301	HM1302	HM1303	HM1304	HM1305	HM1306	HM1307	HM1308	HM1309	HM1310	HM1311	HM1312	HM1313	HM1314	HM1315	HM1316	HM1317	HM1318	HM1319	HM1320	HM1321	HM1322	HM1323	HM1324	HM1325	HM1326	HM1327	HM1328	HM1329	HM1330	HM1331	HM1332	HM1333	HM1334	HM1335	HM1336	HM1337	HM1338	HM1339	HM1340	HM1341	HM134
---	------	----------	----------	-----	------	-----	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	-------

Data Preprocessing

The dataset that is obtained consists of several attributes of each season. Some of those features are less significant or rather nonessential for foreseeing the outcome. So data cleaning is taken out for keeping only those features that are suitable for the forecast. We have also considered converting categorical data to dummy variables as per requirement.

	FTR	HTP	ATP	HM1	HM2	HM3	AM1	AM2	AM3	HTGD	ATGD	DiffFormPts	DiffLP
30	H	1.25	1.00	D	D	W	D	W	L	0.50	0.25	0.25	-16.0
31	NH	0.75	0.25	L	L	W	D	L	L	-0.50	-0.75	0.50	-2.0
32	H	1.00	1.00	L	D	W	D	W	L	0.00	0.25	0.00	-3.0
33	NH	0.75	0.50	L	L	W	D	L	D	-0.25	-0.25	0.25	3.0
34	NH	1.00	1.50	D	L	W	W	W	L	0.00	0.75	-0.50	3.0

Figure 2.Final Dataset after Pre-Processing

The following parameters are being considered in our model.

- Div = League Division
- Date = Match Date (dd/mm/yy) Home Team = Home Team AwayTeam = Away Team
- FTHG = Full Time Home Team Goals FTAG = Full Time Away Team Goals
- FTR = Full Time Result (H=Home Win, D=Draw, A=Away Win) HTHG = Half Time Home Team Goals
- HTAG = Half Time Away Team Goals
- HTR = Half Time Result (H=Home Win, D=Draw, A=Away Win) HTGD = Home Team Goal Difference
- ATGD = Away Team Goal Difference HTP = Home Team Points
- ATP = Away Team Points HM = Home Match
- AM = Away Match

Exploratory Analysis

In our initial exploratory analysis we found out that the home team clearly has an upper edge over the away team.

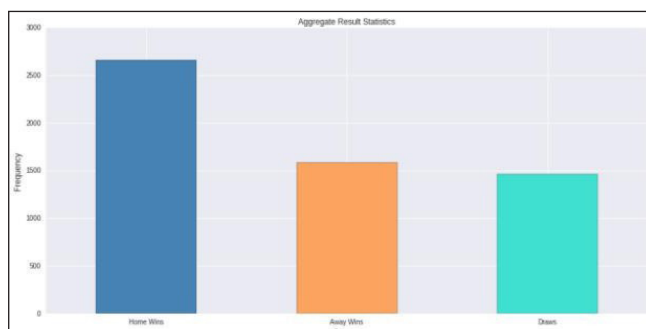


Figure 3.Distribution of Match Wins from 2000-2016

Further it was observed that head to head record alone is not a very reliable factor in predicting the end result of

a soccer game. Combining other factors such as league points and half time result do make a meaningful impact on end result.

Modelling

In this system, we have implemented the following three algorithms: XG Boost, Logistic Regression, Support Vector Machine (SVM).

Logistic Regression: Logistic Regression is a Machine Learning method that is used to solve classification issues. It is a predictive analytic technique that is based on the probability idea. The classification algorithm Logistic Regression is used to predict the likelihood of a categorical dependent variable. The dependent variable in logistic regression is a binary variable with data coded as 1 (yes, True, normal, success, etc.) or 0 (no, False, abnormal, failure, etc.).

Support Vector Machine (SVM): Support Vector Machines are Machine Learning models which are useful for regression analysis and classification tasks. It falls under the supervised learning category of Machine Learning. These are widely used in classification tasks. Support Vector Machines are based on the idea of finding the best hyperplane that divides the dataset into two parts.

XG Boost: The XG Boost stands for extreme Gradient Boosting, which is a boosting algorithm based on gradient boosted decision trees algorithm. XG Boost applies a better regularization technique to reduce overfitting, it is one of the differences from gradient boosting. The 'xg boost' is an open-source library that provides machine learning algorithms under gradient boosting methods. The boost. XGB Classifier is a sci-kit-learn API compatible class for classification.

Experiment

An investigation is performed for obtaining the best precision. In this paper, we are operating the data from past recent seasons of the English Premier League. It is done to determine whether the amount of training data has any impact on forecast precision. Following are the accuracy of each model.

Logistic Regression: F1 score and accuracy for training set: 0.6246 and 0.6654 F1 score and accuracy for test set: 0.6957 and 0.7200.

Support Vector Machine (SVM): F1 score and accuracy for training set: 0.6470 and 0.6978 F1 score and accuracy for test set: 0.7234 and 0.7400.

XGB Classifier: F1 score and accuracy for training set: 0.6470 and 0.6978 F1 score and accuracy for test set: 0.7234 and 0.7400.

XGB Classifier with Grid Search CV: F1 score and accuracy

for training set: 0.6318 and 0.6777 F1 score and accuracy for test set: 0.7234 and 0.7400.

The results were not only less accurate but they also showed overfitting as we can clearly observe from the accuracy score of training and testing data. In order to get better results ensemble model was required. We studied various other literatures and work and found out that ensemble techniques have been used up in the past giving out an accuracy of about 75%.

Since we are already using Xg boost which is an ensemble machine learning algorithm based on gradient boosting, the hyperparameters were tuned by experimenting. The accuracy clearly increased but the overfitting was not resolved. Keeping in mind the variance bias tradeoff, we achieved an accuracy of about 76%. Also our model is more reliable as we have used 5 fold cross-validation on the same.

Result Analysis

Our examination is to foresee the result of the match and when integrated with training data, the XG boost combined with gridsearch cv not only gave better accuracy but also the model is more reliable due to 5 fold cross-validation. So, we have formalized the dataset during the pre-processing phase. Normalization is done to bring the attributes of the training dataset to the same scale. The goal of normalization is to vary the values of numeric columns in the dataset to an ordinary scale, without warping the disparities in the ranges of values.¹¹ Our model is achieving a f1 score of 76.05% and accuracy of 78.59% on training data and f1 score of 75.00% and accuracy of 76.00% on testing data.

Conclusion

Sports Analytics is a rapidly growing field and with the advancement of machine learning algorithms, machine learning can be utilized in this domain also. Our aim was to create a model which could effectively predict the result of a soccer game which can then be utilized in various fields like performance analysis, betting industry and fantasy leagues. The game of football not only depends on numbers but also a lot on players and other factors. Also football is a unpredictable sport and combined with the fact that games are usually low scoring, expecting a very high accuracy is not possible. This research can further be improved by taking into account other factors such as a players health statistic or sentiment analysis from twitter.

References

1. Thamaraimanalan T, Naveena D, Ramya M et al. Prediction and classification of fouls in soccer game using deep learning. *Irish interdisciplinary Journal of science & research* 2020; 4(3): 66-78.
2. Yang. Predict soccer match outcome based on player performance, *Francis Academic Press*, UK, 2021; 3(3): 74-78. ISSN 2618-1576.DOI: 10.25236/fsr.
3. Martinovich J, Snásel V, Ochodkova E et al. Robot soccer-strategy description and game analysis. in *ecms* 2010; 265-270.
4. Matej U, Gustav Š, Ondřej H et al. Optimal Sports Betting Strategies In Practice: An Experimental Review, *IMA Journal Of Management Mathematics* 2021. 10.1093/IMAMAN/DPAA029.
5. Lee WJ, Jhang HJ, Choi SH. Fuzzy Study On The Winning Rate Of Football Game Betting. *Advances In Technology Innovation* 2021; 6(3): 169-178. 10.46604/AITI.2021.6517
6. Wheatcroft E. Forecasting Football matches by predicting match statistics. *Journal Of Sports Analytics* 2021; 1-21. 10.3233/JSA-200462
7. Baboota R, Kaur H. Predictive Analysis And Modeling Football Results Using Machine Learning Approach For English Premier League. *International Journal of Forecasting* 2019; 35(2); 741-755. 10.1016/J.IJFORECAST.2018.01.003
8. Rahman MM, Shamim F, MO et al. An Analysis of Bangladesh One Day International Cricket Data: A Machine Learning Approach. *International Conference on Innovations in Science, Engineering and Technology (ICISSET)* 2018.
9. Football-data. (2019). data | football-data, [online] Available at: <http://www.football-data.co.uk/> [Accessed on 27 Oct. 2021].
10. Oughali MS, Bahloul M, El Rahman SA. Analysis of NBA Players and Shot Prediction Using Random Forest and XGBoost Models. *International Conference on Computer and Information Sciences (ICCIS)* 2019.
11. Anik AI, Yeaser S, Hossain AGMI et al. Player's Performance Prediction in ODI Cricket Using Machine Learning Algorithms. *4th International Conference on Electrical 33 Engineering and Information & Communication Technology(iCEEICT)* 2018.
12. Thirumalai C, Kanimozhi R, Vaishnavi B et al. Data analysis using box plot on electricity consumption. *International Conference of Electronics, Communication and Aerospace Technology(ICECA)* 2017.
13. Statistics how to data science central Normalized | statisticshowto.datasciencecentral. [online] Available at 2015: <https://www.statisticshowto.datasciencecentral.com/normalized/> [Accessed on 27 Oct. 2021].
14. Like geeks. Seaborn-heatmap-tutorial | likegeeks. [online] Available at: 2019 <https://likegeeks.com/seaborn-heatmap-tutorial/> [Accessed on 15 Jan 2022].
15. Towards data science. Hyper parameter 1Tuningc5619e7e6624 | towards data science. [online] Available at: 2019 <https://towardsdatascience.com/hyperparameter-tuning1c5619e7e6624> [Accessed on 27 Oct. 2021]