A Model for Predicting Music Popularity on Spotify

Carlos Vicente Soares Araujo

Data Scientist – Loggi

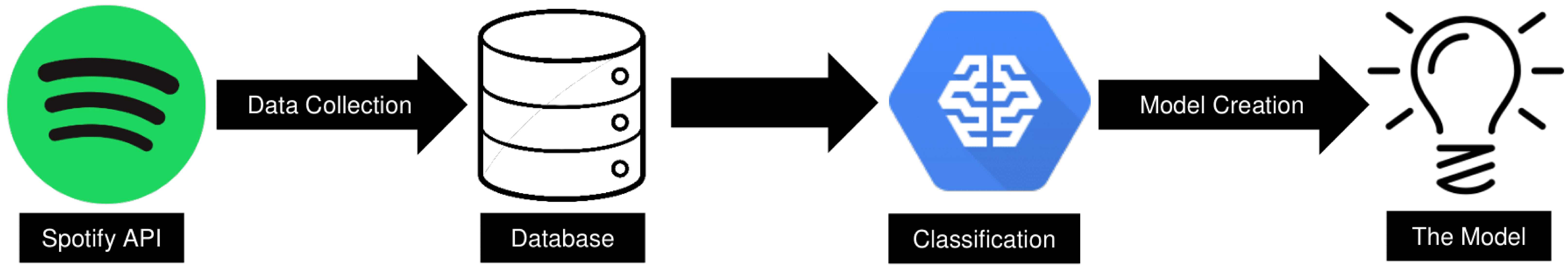
INTRODUCTION

The way people listen to music is changing. In 2018, for the first time, streaming became the main form of music consumption, accounting for 47% of the music market, according to the International Federation of the Phonographic Industry (IFPI) annual report. In 2019 this percentage was even higher accounting for 56.1% of global music revenues. Therefore, streaming has become critical for artists to achieve good business results.

One way to help artists and record labels maximize commercial return is to use a model to predict whether their music will be popular on streaming platforms. A prediction model could give artists and labels an edge over competitors, because they could focus more on songs that tend to earn a good yield.

Here, I present a model to predict if a song will be popular on Spotify streaming platform even before its release. Spotify was chosen as my study case because it is one of the world's largest music streaming service in number of users.

METHODOLOGY



RESULTS

	SVM		Gaussian Naive Bayes		Logistic Regression		KNN	
	PM	ROM	PM	ROM	PM	ROM	PM	ROM
Accuracy	0.8511	0.5330	0.8481	0.5353	0.8403	0.5336	0.8395	0.5433
Precision	0.9650	0.6194	0.8719	0.5485	0.8734	0.5843	0.8947	0.6293
NPV	0.7534	0.4573	0.8161	0.4338	0.7980	0.4329	0.7774	0.4667
Recall	0.7706	0.5002	0.8644	0.6805	0.8467	0.6719	0.8190	0.5123
Specificity	0.9616	0.5781	0.8257	0.3360	0.8315	0.3437	0.8677	0.5858
F1 Score	0.8569	0.5534	0.8681	0.6289	0.8598	0.6250	0.8552	0.5648
AUC	0.8661	0.5391	0.8450	0.5083	0.8391	0.5078	0.8433	0.5491
MCC	0.7253	0.0775	0.6890	0.0174	0.6748	0.0164	0.6793	0.0971

Table 1: Performance of the models for the experiment where the predictions were made by day.

	SVM		Gaussian Naive Bayes		Logistic Regression		KNN	
	PM	ROM	PM	ROM	PM	ROM	PM	ROM
Accuracy	0.9081	0.5882	0.8456	0.6324	0.8235	0.6838	0.8713	0.6360
Precision	0.9130	0.3707	0.7381	0.3200	0.6667	0.3333	0.9273	0.3651
NPV	0.9064	0.7500	0.8936	0.7027	0.9176	0.7000	0.8571	0.7177
Recall	0.7683	0.5244	0.7561	0.1951	0.8293	0.0488	0.6220	0.2805
Specificity	0.9684	0.6158	0.8842	0.8211	0.8211	0.9579	0.9789	0.7895
F1 Score	0.8344	0.4343	0.7470	0.2424	0.7391	0.0851	0.7445	0.3172
AUC	0.8684	0.5701	0.8603	0.5081	0.8560	0.5033	0.8004	0.5350
MCC	0.7770	0.1301	0.6360	0.0192	0.6164	0.0149	0.6866	0.0761

Table 2: Performance of the models for the experiment where the predictions were made by day.

	Experiment 1	Experiment 2
Accuracy	0.9081	0.5882
Precision	0.9130	0.3707
NPV	0.9064	0.7500
Recall	0.7683	0.5244
Specificity	0.9684	0.6158
F1 Score	0.8344	0.4343
AUC	0.8684	0.5701
MCC	0.7770	0.1301

Table 3: Higher performance percentages achieved by PM over ROM.

REFERENCES

[1] D. Herremans, D. Martens, and K. Sörensen, "Dance hit song prediction," Journal of New Music Research, vol. 43, no. 3, pp. 291–302, 2014. [Online]. Available: https://doi.org/10.1080/09298215.2014.881888

[2] M. Reiman and P. Örnell, "Predicting hit songs with machine learning," ser. TRITA-EECS-EX, no. 2018:202, 2018.

[3] C. V. Araujo, R. M. Neto, F. G. Nakamura, and E. F. Nakamura, "Predicting music success based on users' comments on online social networks," in Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, ser. WebMedia '17. New York, NY, USA: ACM, 2017, pp. 149–156. [Online]. Available: http://doi.acm.org/10.1145/3126858.3126885

Related papers: https://scholar.google.com.br/citations?user=0kZvtr4AAAAJ
Contact: vicente@icomp.ufam.edu.br