

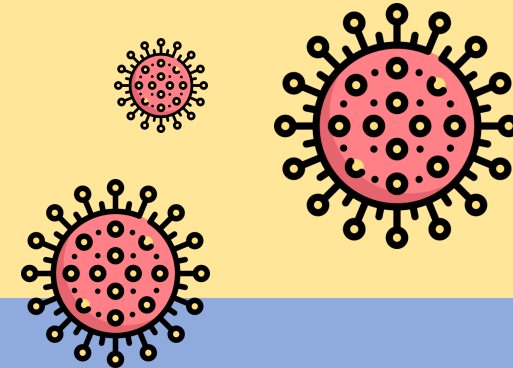


COVID-19 Community Mobility Reports

Unidade Curricular: Inteligência Artificial

Maria Barros (up201608444) e Miguel Ferreira (up201606158)

Definição do problema



Antes da definição do problema, é necessário primeiro conhecer o dataset. Para este trabalho, o dataset com que vamos trabalhar possui informação sobre **19 diferentes países**, com informações diárias durante **43 dias**, sobre diferentes informações:



Tendências de mobilidade a locais como **supermercados, farmácias...**



Tendências de mobilidade a locais como **restaurantes, shoppings...**



Tendências de mobilidade a **locais de trabalho**



Tendências de mobilidade a locais como **parques, jardins, marinas...**



Tendências de mobilidade a **transportes públicos**



Tendências de mobilidade a **locais de residência**



Número total de casos

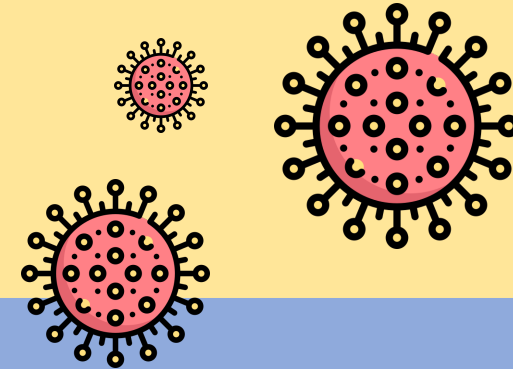


Número total de mortes

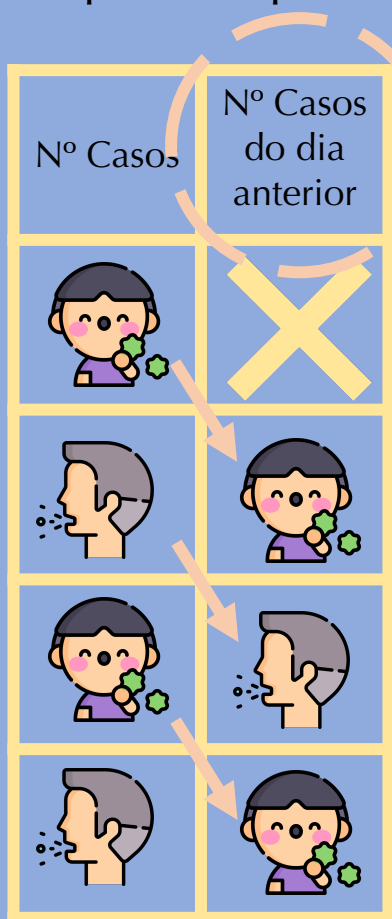


Como primeiro problema, decidimos então focar-nos na **previsão do número total de casos**, utilizando como **features todas as tendências de mobilidade**, e ainda o **número de casos do dia anterior**

Ferramentas a utilizar



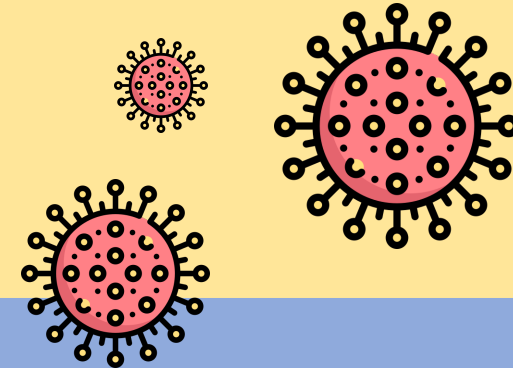
A primeira parte do trabalho consistiu então numa primeira avaliação do dataset



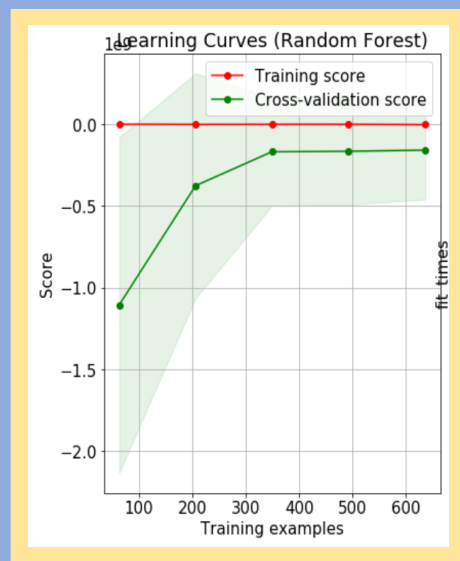
O primeiro passos consistiu então na criação de uma nova coluna, em que cada entrada corresponde ao número de **casos no dia anterior**. Desta vez, para prever o número de casos para cada dia, este valor pode ser utilizado como característica, e ajudar nessa previsão

De forma a dificultar a previsão, considerámos, noutra abordagem, não o **número de casos do dia anterior** (dado que facilita bastante a tarefa da previsão), mas sim o número de casos conhecidos **na semana anterior**

Ferramentas a utilizar



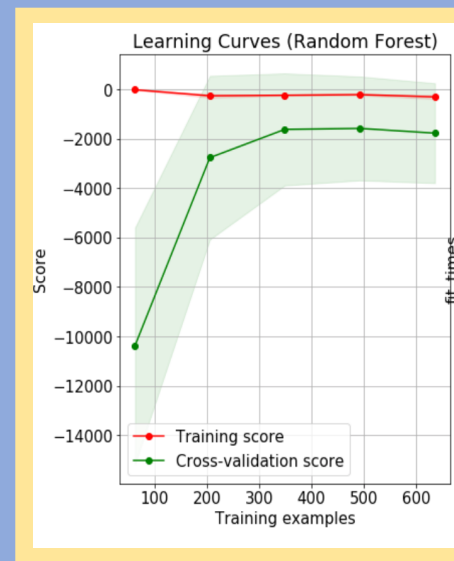
De seguida, e de forma a conseguir averiguar quais os algoritmos que produziriam melhores resultados para o dataset, traçaram-se as **curvas de aprendizagem** para diferentes algoritmos. Os algoritmos testados foram **Regressão Linear, KNN, Naïve Bayes, SGD, SVC, MLP, Decicion Tree e Random Forest**, estando os resultados da linha de aprendizagem apresentados a seguir. Várias métricas foram utilizadas para avaliar os resultados



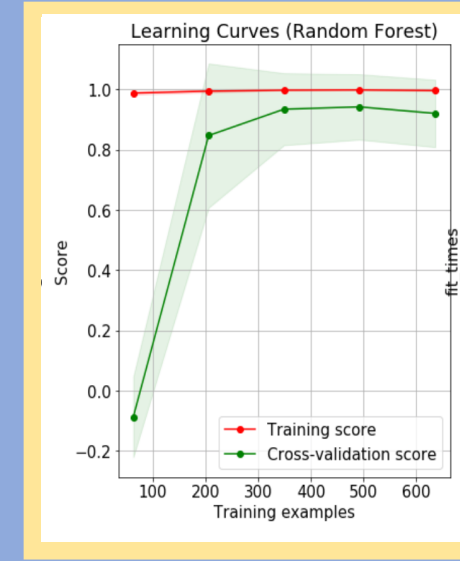
Negative Mean Squared Error



Max Error

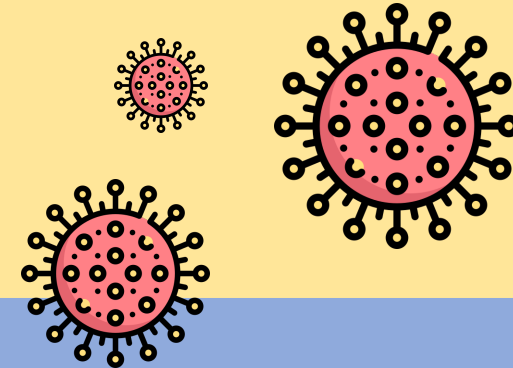


Negative Mean Absolute Error



R²

Resultados



R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
0,99878	0,99676	1310552,8	0,00002	18265,814	0,05078	245,7842	0,00194

TREINO

R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
0,93476	0,90152	164025619	0,00246	54180,342	0,18756	1682,94496	0,01134

TESTE

R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
0,99378	0,98378	8064909,04	0,0002	38179,576	0,1099	802,24696	0,0054

TREINO

R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
-4,77152	0,52556	440367201	0,02088	77400,364	0,42828	7120,1995	0,07456

TESTE

R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
0,97032	0,92196	36840099,2	0,00098	88155,719	0,2017	1687,35336	0,01384

TREINO

R2	R2_s	MSE	MSE_s	ME	ME_s	MAE	MAE_s
-3,82542	-0,0736	446198936	0,03566	80631,1427	0,73512	6296,10446	0,09896

TESTE

PREVISÃO DO NÚMERO DE CASOS A PARTIR DO NÚMERO DE CASOS DO DIA ANTERIOR

Resultados muito satisfatórios tanto no **set de treino** e no **set de teste**

PREVISÃO DO NÚMERO DE CASOS A PARTIR DO NÚMERO DE CASOS DA SEMANA ANTERIOR

Resultados muito satisfatórios tanto no **set de treino** mas muito mais no **set de teste**

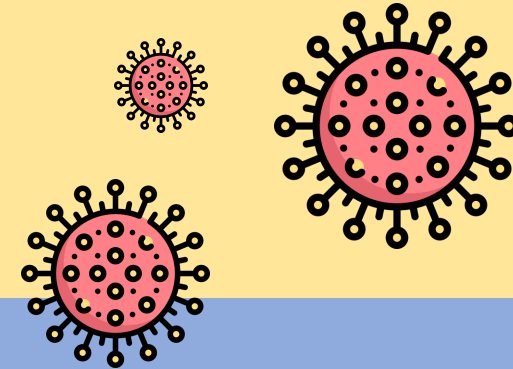


Indica **overfitting**

TUNNING DOS HYPERPARÂMETROS DO MODELO ANTERIOR

Melhoramento dos resultados, mas ainda existe bastante **overfitting**

Trabalho futuro



De forma a tentar reduzir o overfitting dos modelos



Utilizar técnicas de **feature selection**
(como **PCA**, por exemplo)

Testar outras problemáticas com o dataset disponível



Tentar prever o **número de fatalidades**