CarsInLoc

Problem statement

This data product is aimed to car-sharing companies that want to **predict how many cars** they will have in a **particular location** in a precise moment in time.

What are the prerequisites?

The company needs to provide a register with the following information: location (**latitude** and **longitude**), **number of cars** in that location and **timestamps**.

timestamp	latitude	longitude	total_cars	carsList
2019-01-10 11:45:...	32.09995	34.78794	1	[182]
2019-01-10 11:45:...	32.06567	34.79612	1	[268]
2019-01-10 11:45:...	32.06465	34.80322	1	[106]
2019-01-10 11:45:...	32.05978	34.81034	1	[180]
2019-01-10 11:45:...	32.05133	34.75089	1	[16]
2019-01-10 11:45:...	32.04223	34.7742	1	[72]
2019-01-10 11:45:...	32.04156	34.77128	1	[160]
2019-01-10 11:45:...	32.12373	34.81346	1	[210]
2019-01-10 11:45:...	32.11874	34.83406	1	[136]
2019-01-10 11:45:...	32.03351	34.75509	1	[27]
2019-01-10 11:45:...	32.14288	34.79361	1	[75]
2019-01-10 11:45:...	32.14306	34.79729	1	[132]
2019-01-10 11:45:...	32.083175	34.776552	0	[]
2019-01-10 11:45:...	32.088379	34.775111	0	[]
2019-01-10 11:45:...	32.074877	34.773515	0	[]
2019-01-10 11:45:...	32.098603	34.778565	0	[]
2019-01-10 11:45:...	32.09478	34.79728	0	[]
2019-01-10 11:45:...	32.098032	34.798089	0	[]
2019-01-10 11:45:...	32.12047	34.800318	0	[]
2019-01-10 11:45:...	32.04409	34.80421	0	[]

Algorithms

A Machine Learning algorithm based on **Apache Spark ML** (a novel library that provides **parallel computing**) and a Deep Learning Algorithm based on **Keras** (one of the most used DL libraries) have been developed and tested. The **accuracy** of the predictions is around **73%**.