

Project overview

Introduction: Luckin Coffee is a Chinese company which was listed on Nasdaq on May 16th, 2019. Early this year, Muddy Waters revealed a report which claimed that Luckin Coffee was inflating its number of items sold per store. After Luckin announced on April 2 that it would conduct an internal investigation into fraud allegations, Luckin's stock price dropped more than 80%. This event aroused our interests. For this project, we decided to use Luckin's earning announcements to develop a graphical database to apply the model to autocomplete/word completion.

Approach: For mid-term submission, we focused on the python. For final submission, we will integrate with neo4j.

Methodology:

NLTK: Python's build-in programs to work with language data. It provides us with sample texts and tutorial code.

N-gram model: the model that can assign probabilities to the sequences of words. It allows our program to output the predicted text.

Neo4j: the graph database software. It can help us to build graph database which can be connected to python programs by API.

Mid-term output: All works have been uploaded to this folder. A python program allows us to use n-gram model to predict text. For the remaining work, we will integrate with the graphical database created by neo4j.