# Application of extracting raw data from the internet.

Ex: Analyze user feedback about phone product

Algorithm: Logistic regression

Concept model: Keyword from user -> Get data -> Product report + Compare

Objective: user -> keyword -> report

Crawl data: Remove all the HTML code in page to get raw data

Pre-processing for data: remove stop word (ex: negative word), bag of words, TF-IDF (thuat toán trong so)

Bag of words: is the one-way matrix include n elements which are get from separating word of data -> Often not exactly

TF-IDF: -> Find the importance of the specific word with its position -> Improve the drawback of bag of words.

Model of application: User -> Brain –(Request data)-> Database -> Brain (-(Question)-> Workers –(Get data)-> Database -> Brain)

Crawl data: Selenium web-driver + Chrome Driver

* Gia lap de su dung automating texting in order to get raw data from website.

Gia lap nut “end” de load comment cho youtube vi that platform phai check dung browser moi cho load comment.

Bat buoc phai pre-processing data