Team Name : SKYNET

Team Members :

* Salamat Burzhuev
* Muhammad Fauzi Sahdan
* Liew Nam Yang

Our prediction model is ensemble of 4 different models. We train each groups of categories (beauty,fashion and mobile ) separately.

Here is our summary:

* Model: Logistic regression; Tokenization: Countvectorizer:
* Beauty parameters :
* CountVectorizer(ngram\_range=(1,3))),('model',LogisticRegression(C=0.8)
* Fashion parameters
* CountVectorizer(stop\_words =ignore\_words,ngram\_range=(1,5))),('model',LogisticRegression(C=0.4))])
* Mobile parameters
* CountVectorizer(ngram\_range=(1,3),min\_df=2)),('model',LogisticRegression(C=1)
* Model: Naïve Bayes(MultinomialNB); Tokenization: Countvectorizer
* Beauty
* ', CountVectorizer(stop\_words =ignore\_words,ngram\_range=(2,6))),('model',nb)]
* Mobile: CountVectorizer(stop\_words =ignore\_words,ngram\_range=(1,2))),('model',nb)])
* Fashion:', CountVectorizer(stop\_words =ignore\_words,ngram\_range=(2,6))),('model',nb)])
* Model: word2vec. Created word2vec based on our train and test titles size 100
* pipe\_beauty=Pipeline([('model',RandomForestClassifier(n\_estimators=280))])
* pipe\_mobile=Pipeline([('model',RandomForestClassifier(n\_estimators=200))]) # I just choose lowest complexity
* pipe\_fashion=Pipeline([('model',RandomForestClassifier(n\_estimators=300))])
* Image classification: Model: resnet34,
* transfer learning

data augumentation

learning rate anneling

using lower rates for training internal layers

Countvectorizer gridsearch

Logistic Regression

Word2vec