

Stack Overflow Tag Recommendation System using One-vs-Rest Multi-label Classification

PRESENTED BY: TRIXIA BELLEZA

ADVISER: CONCEPCION KHAN

Background of the Study

STACK OVERFLOW

- Q & A Forum for programmers
- Questions are organized in a form of tags

TAGS

- Describes question content
- For filter-searching of questions

STATEMENT OF THE PROBLEM



Tagging of questions are done manually by users.



Users find it difficult to identify which category their question belongs to

STATEMENT OF THE PROBLEM



Prone to human error



Disarray of information that
could overturn its purpose

SIGNIFICANCE OF THE STUDY

**TO LESSEN THE BURDEN
OF MANUAL TAGGING BY
PROVIDING ACCURATE
TAGS**

**AN AUTOMATED TEXT-
CLASSIFICATION IS
NEEDED**

OBJECTIVES OF THE STUDY



Generally, the study aims to create an application that recommends tags to Stack Overflow questions using a multi-label classifier.

OBJECTIVES OF THE STUDY

01

To gather 1K training and testing data from Stack Overflow using its API

02

To create a multi-label classifier (OVR) using a Support Vector Machine

OBJECTIVES OF THE STUDY

03

To test the accuracy of the machine (60% threshold) by computing precision, recall, and F-score.

04

To create a chrome extension that recommends tags to Stack Overflow questions

Review of Related Literature

GUAN ET AL (2009)

- The focus of their study lies on the documents that are annotated by users.

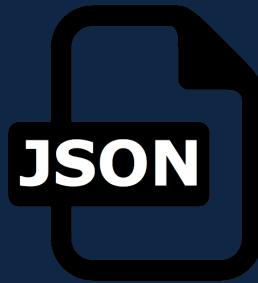
- Collaborative filtering

**KRESTEL &
FRANKHAUSTER**

- Tag recommendation is based on tagging history of the user and textual content.



Python

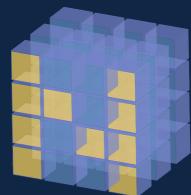


JSON



Scikit Learn

METHODOLOGY



NumPy

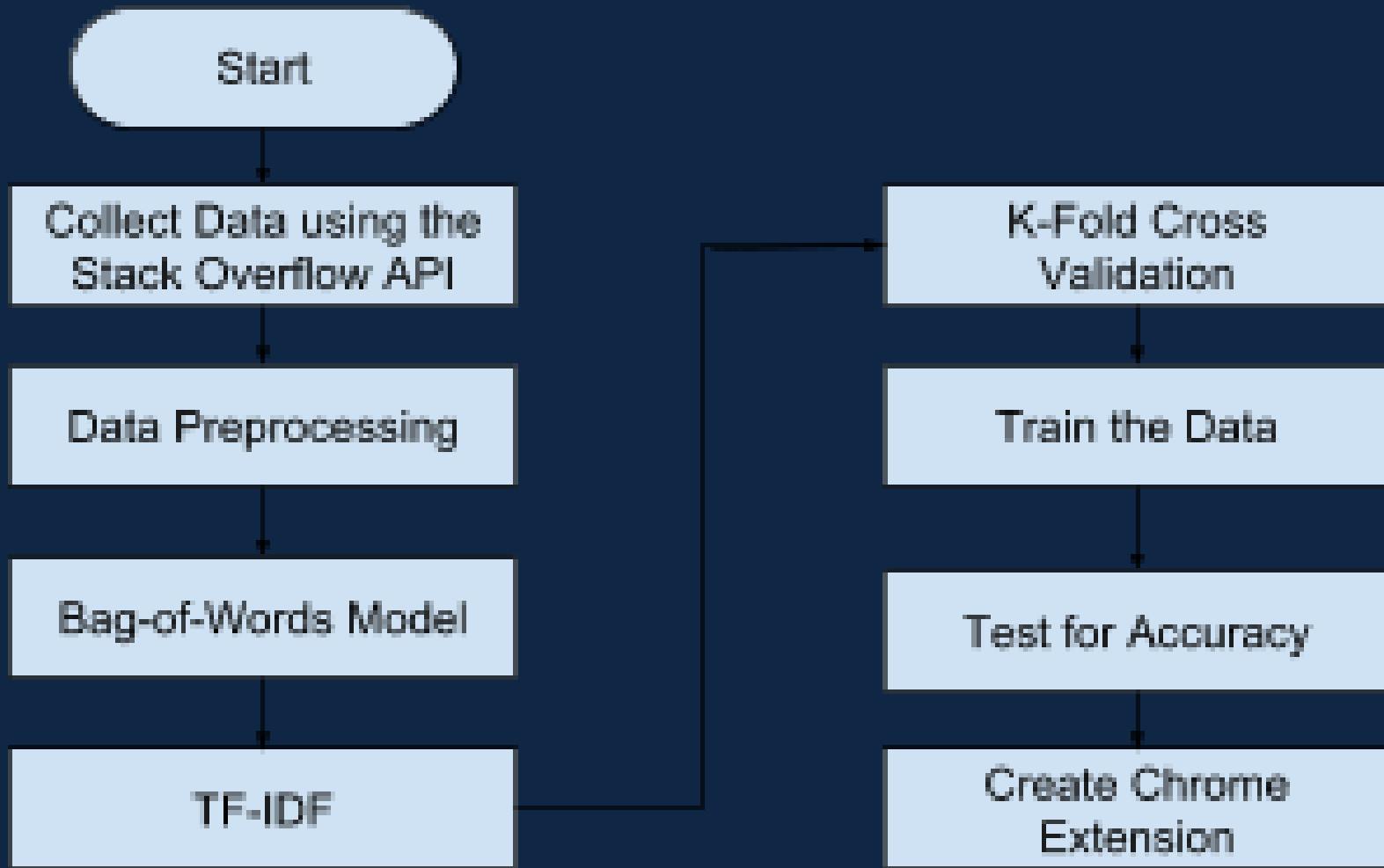


Numpy



NPM

Rapydscript



DATA COLLECTION

Stack Overflow provides
an API that outputs a
JSON file

Data Preprocessing using NLTK

01

Ignore stop words

02

**Ignore punctuation
marks**

03

**Convert all words
to lower case**

Actual Data Labels

5 commonly used tags

Java, Python, R, C,
JavaScript

Control Flow Statements

if-then, if-then-else, for-loop,
while-loop,

BAG-OF-WORDS

Word frequency count

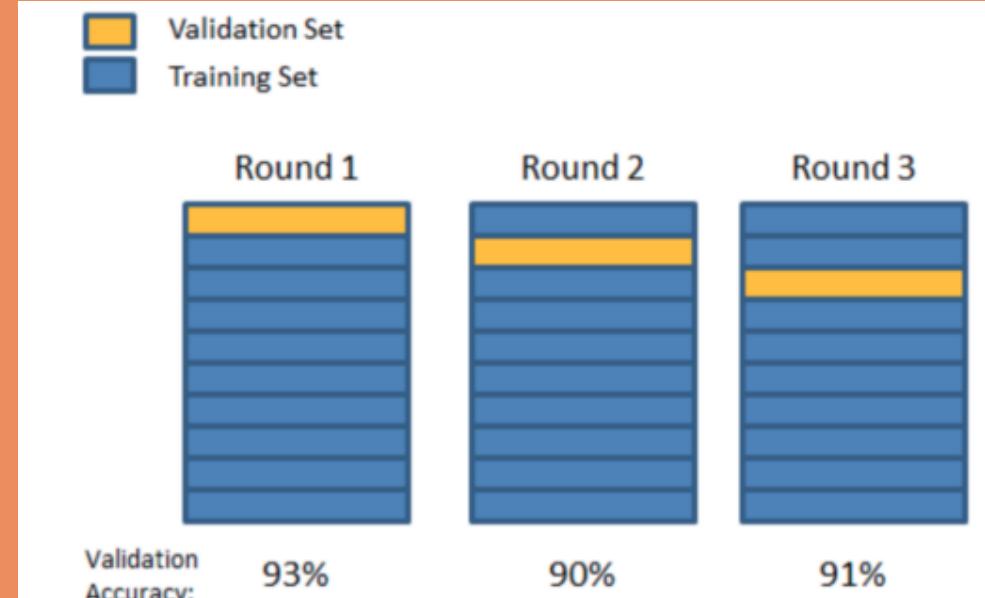
TERM FREQUENCY INVERSE DOCUMENT FREQUENCY

- If a word appears frequently in a document, it's important. Give the word a high score.
- But if a word appears in many documents, it's not a unique identifier. Give the word a low score.

K-Fold Cross Validation

K depends on the number of datasets to be used

Split our data into k different subsets (or folds)



Multi-label Classification

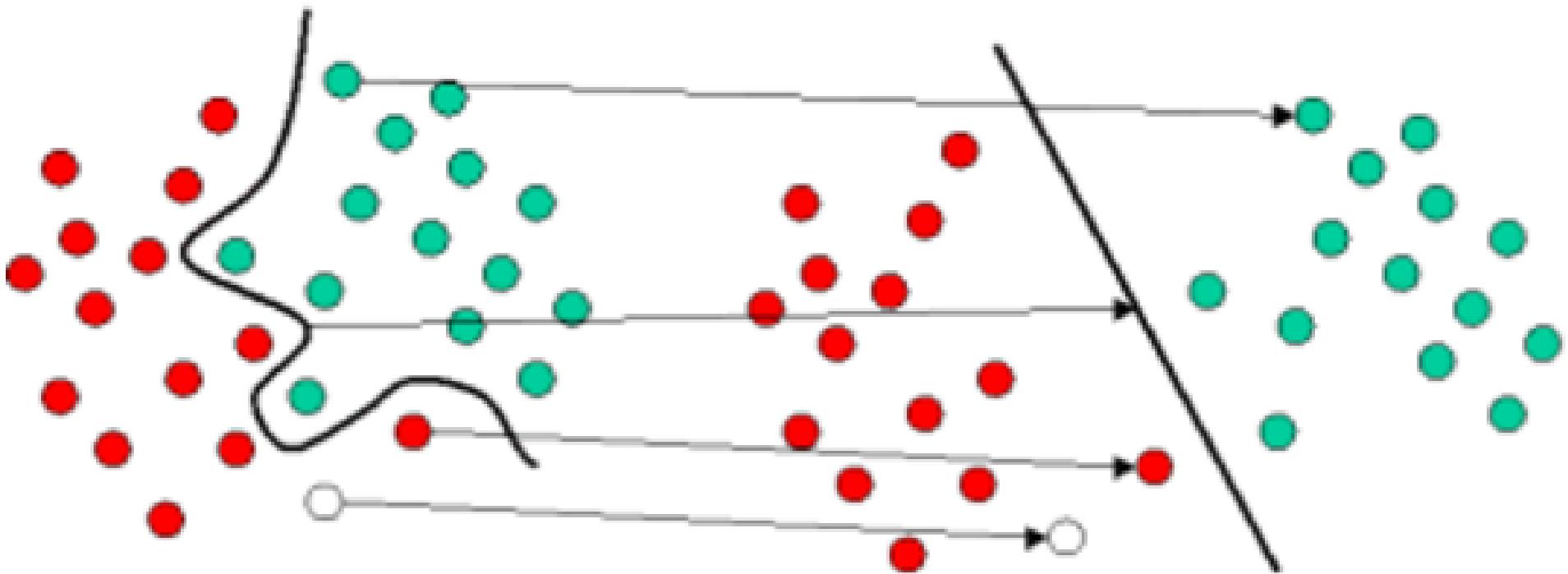
One-vs-Rest

**Support Vector
Machine**

**Training
the model**

Input space

Feature space



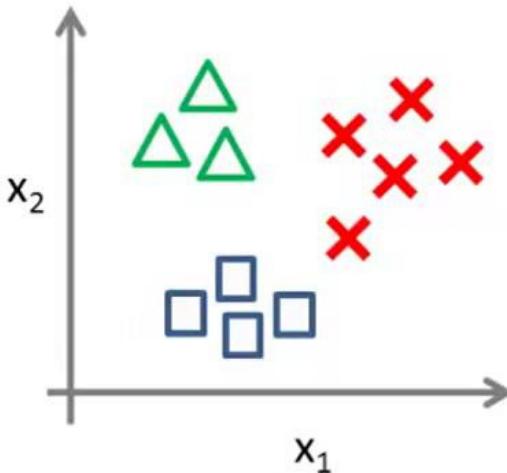
Why SVM?

HERE WE SEE THE ORIGINAL OBJECTS REARRANGED USING KERNELS. SVM IS PARTICULARLY SUITED TO HANDLE SUCH TASKS.

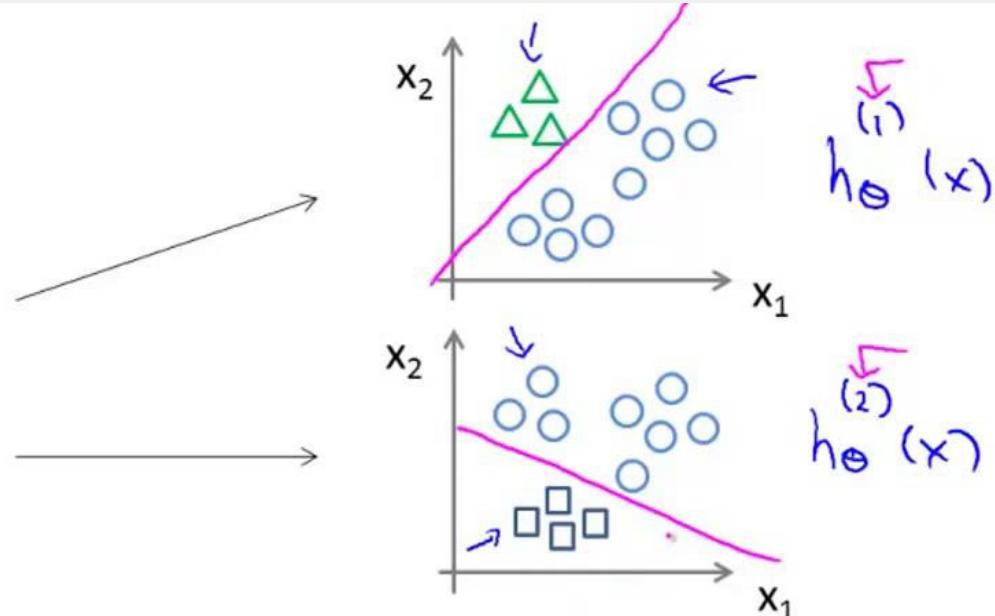
WHY USE ONE-VS-REST?

ONE-VS-REST

One-vs-all (one-vs-rest):



- Class 1: ←
- Class 2: ←
- Class 3: ←



Performance Evaluation

Precision

ratio of correctly classified samples to the total classified samples

Recall

ratio of correctly classified samples to the total samples in the actual class

F-score

determines the accuracy of a classifier

I'm trying to make a function that will compare multiple variables to an integer and output a string of three letters. I was wondering if there was a way to translate this into Python. So say:

```
x = 0  
y = 1  
z = 3  
mylist = []  
if x or y or z == 0 :  
    mylist.append("c")
```

draft saved

I'm trying to make a function that will compare multiple variables to an integer and output a string of three letters. I was wondering if there was a way to translate this into Python. So say:

```
x = 0  
y = 1  
z = 3  
mylist = []  
  
if x or y or z == 0 :  
    mylist.append("c")  
if x or y or z == 1 :  
    mylist.append("d")  
if x or y or z == 2 :  
    mylist.append("e")  
if x or y or z == 3 :  
    mylist.append("f")
```

which would return a list of

["c", "d", "f"] Is something like thi

Tags

e.g. (linux vba .net)

Suggested Tags

Post Your Question

Discard

Suggested Tags

Python

If-statement

Boolean-logic

Select All

Done

How to Ask

Is your question about programming?

We prefer questions that can be answered, not just discussed.

Provide details. Share your research.

If your question is about this website, [ask it on meta](#) instead.

[visit the help center »](#)
[asking help »](#)

