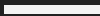Classify a given research paper using Natural Language Processing.

Why solve this problem ?

Given the large archive of scientific articles, the ability to find relevant papers has become difficult.
I provide the ability to filter research papers by field so as to reduce the bane of identifying relevant articles.

CONTENTS

DATA OVERVIEW.

DATA
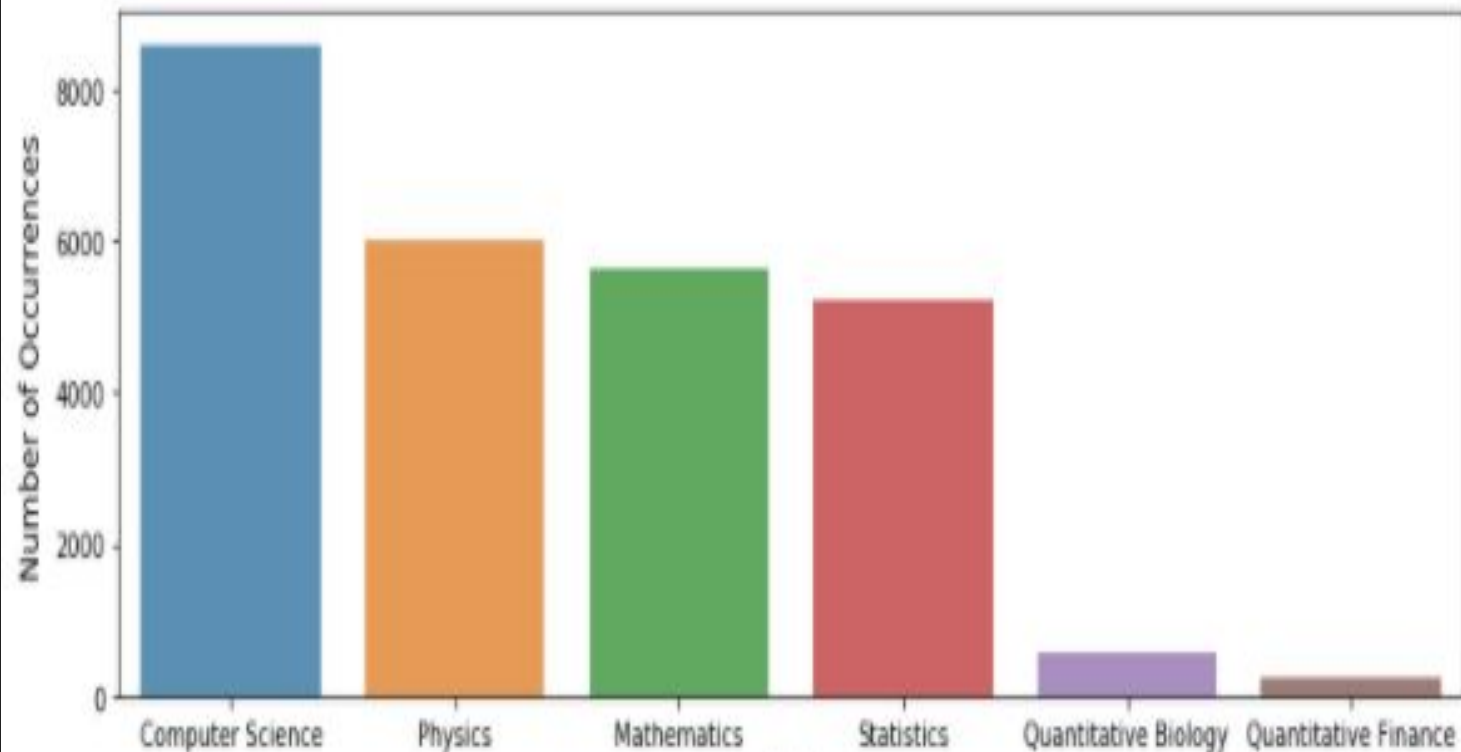
20972 Entries

Title
Abstract
Computer Science
Physics
Mathematics
Statistics
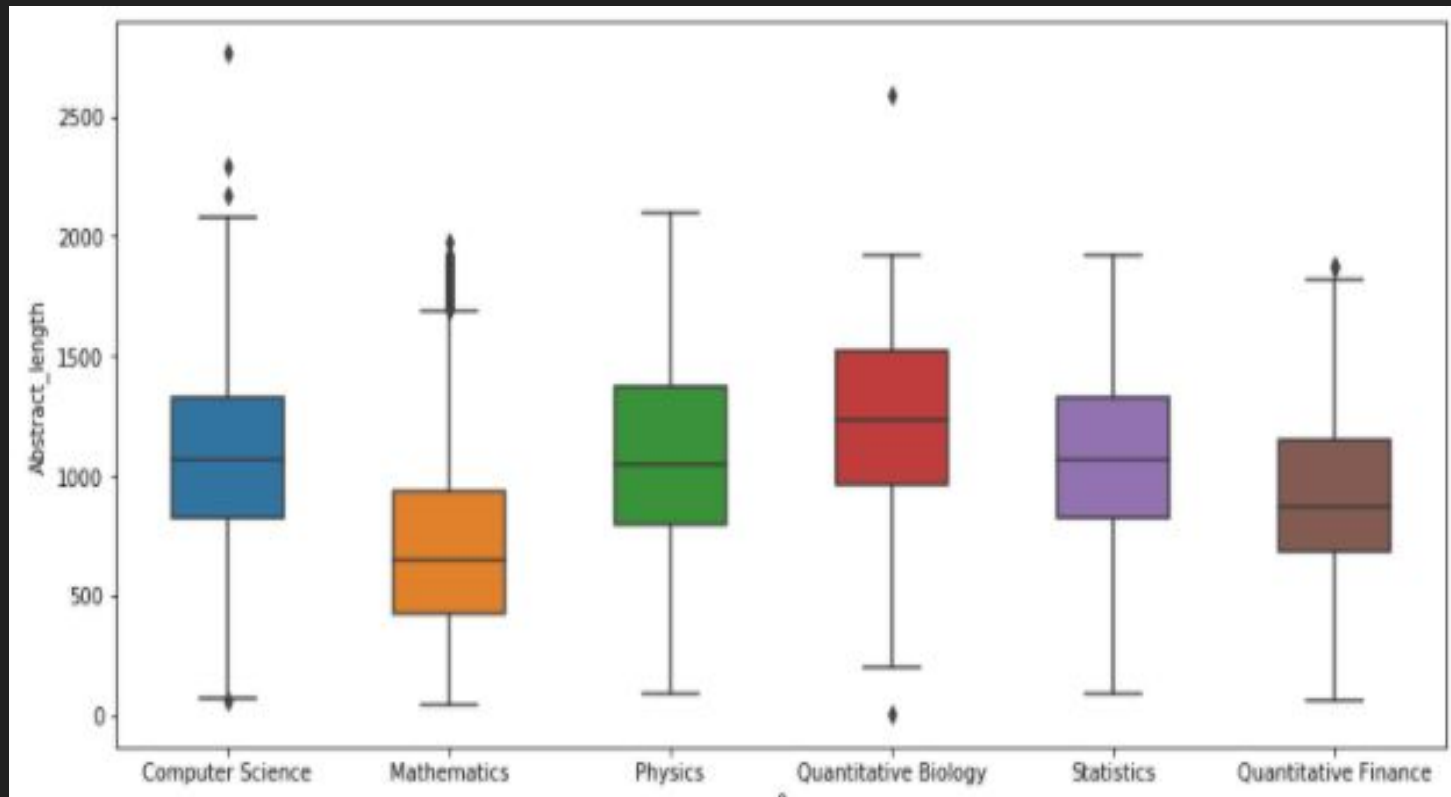Quantitative Finance
Quantitative Biology

INITIAL OBSERVATIONS.

INITIAL OBSERVATIONS.

CHALLENGE

Slow processing power

PROPOSED SOLUTION

STEP 1: ✖ Clean the text

STEP 2: ✖ Convert text into features

STEP 3: ✖ Train different Models

STEP 4: ✖ Compare Models

STEP 5: ✖ Choose the Best Model

Results.

Model	Training Set Accuracy	Test Set Accuracy
Logistic Regression	0.855172	0.819187
KNN	0.808218	0.812512
Random Forest	0.999873	0.787526

RECOMMENDATIONS

Include the KNN model on research website to assist scientists to readily filter 1000's of research papers by field.

Improvements

Build a model that can recommend other research papers which cover similar topics based on the title, abstract and research field