



CRAIGSLIST JOB CLASSIFICATION & SCAM DETECTION

MGMT 590 AUD - Final Project

TEAM EXTNOMICS



Shubham
Goyanka



Hrohaan
Malhotra



Naveen
Kumaran



Adithya
Kothari

ABOUT CLIENT

CLASSIFIED ADS

Network of online communities that offer classified advertisements

CATEGORIES

Sections are devoted to jobs, housing, personals, for sale, services etc.

VERY POPULAR

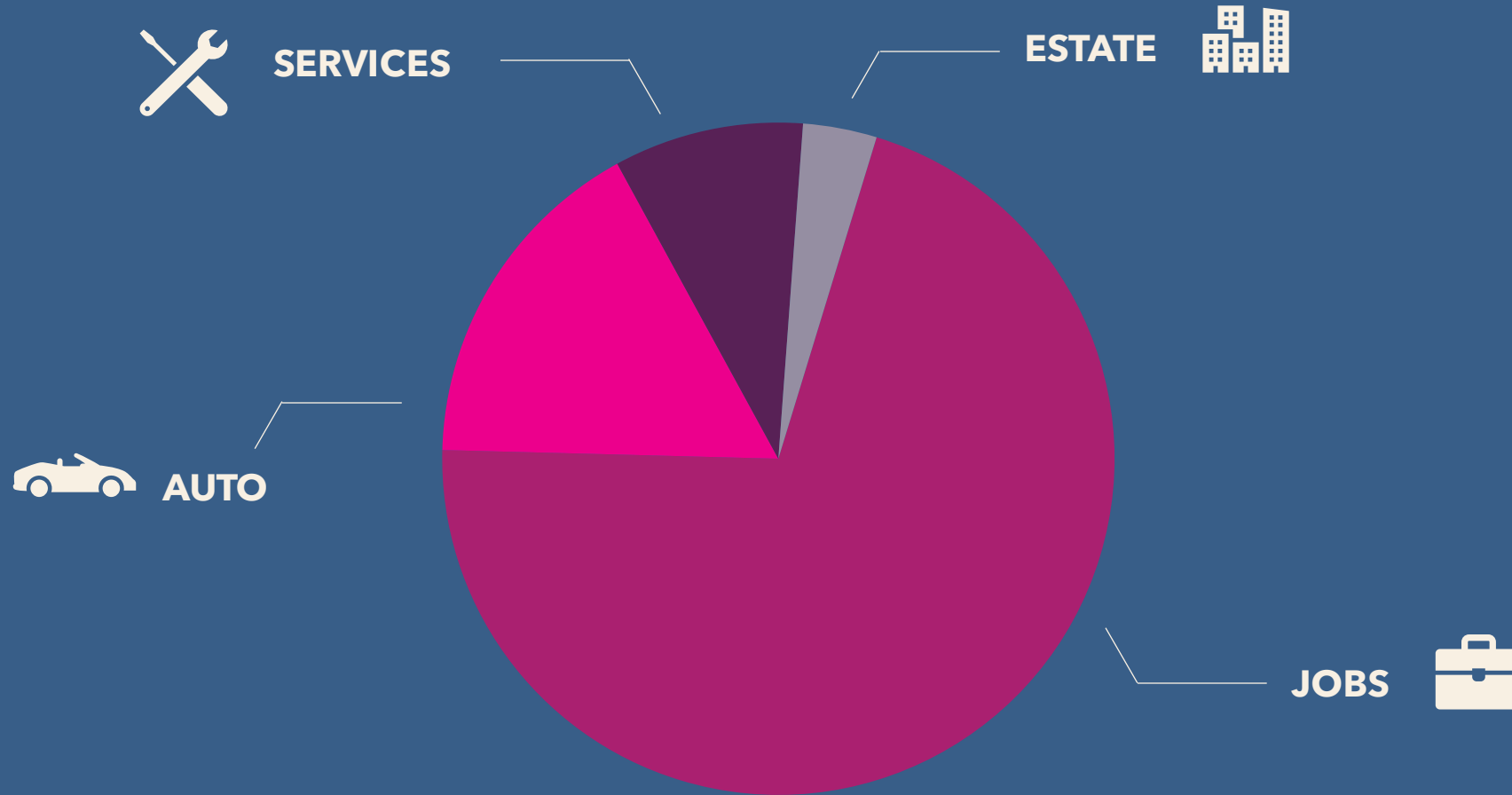
Over 300M visits per month

MISSION

Providing an alternative to impersonal, big-media sites

Restoring the human voice to the Internet, in a humane, non-commercial environment

REVENUE CATEGORIES



REVENUE SOURCES

\$75

● **Job Postings** in Bay area

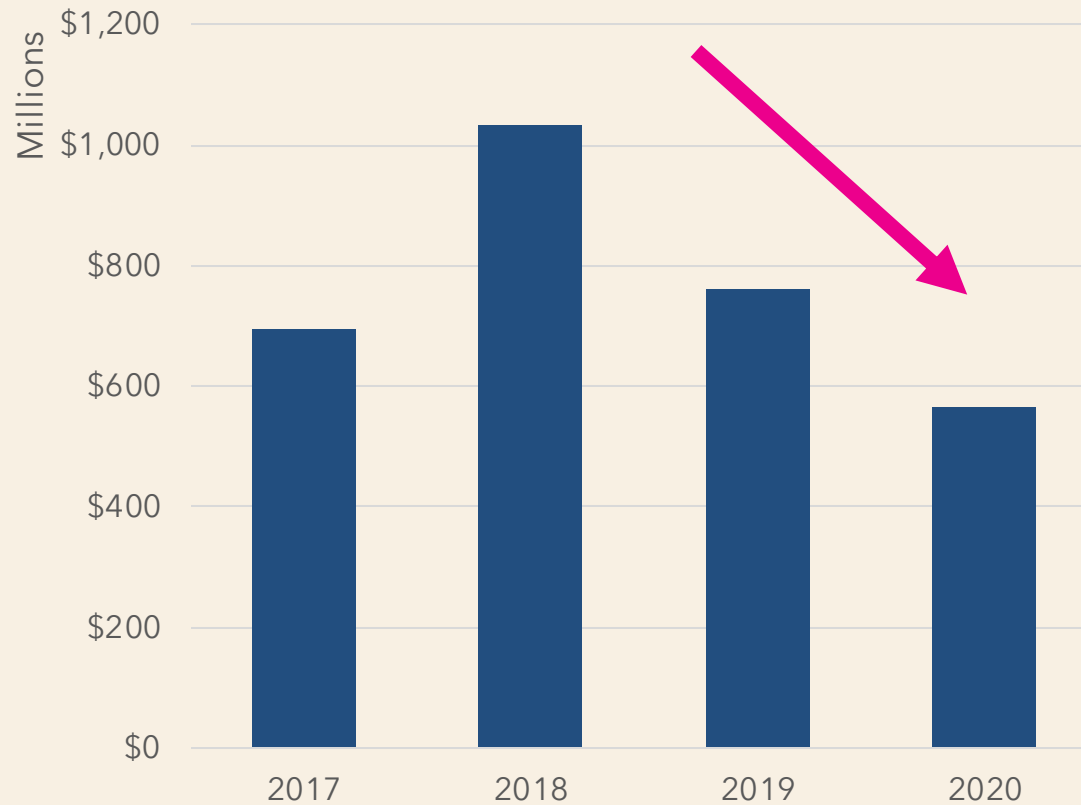
\$25

● **Other Job Postings**

\$10

● **Apartment postings**

YOY REVENUE

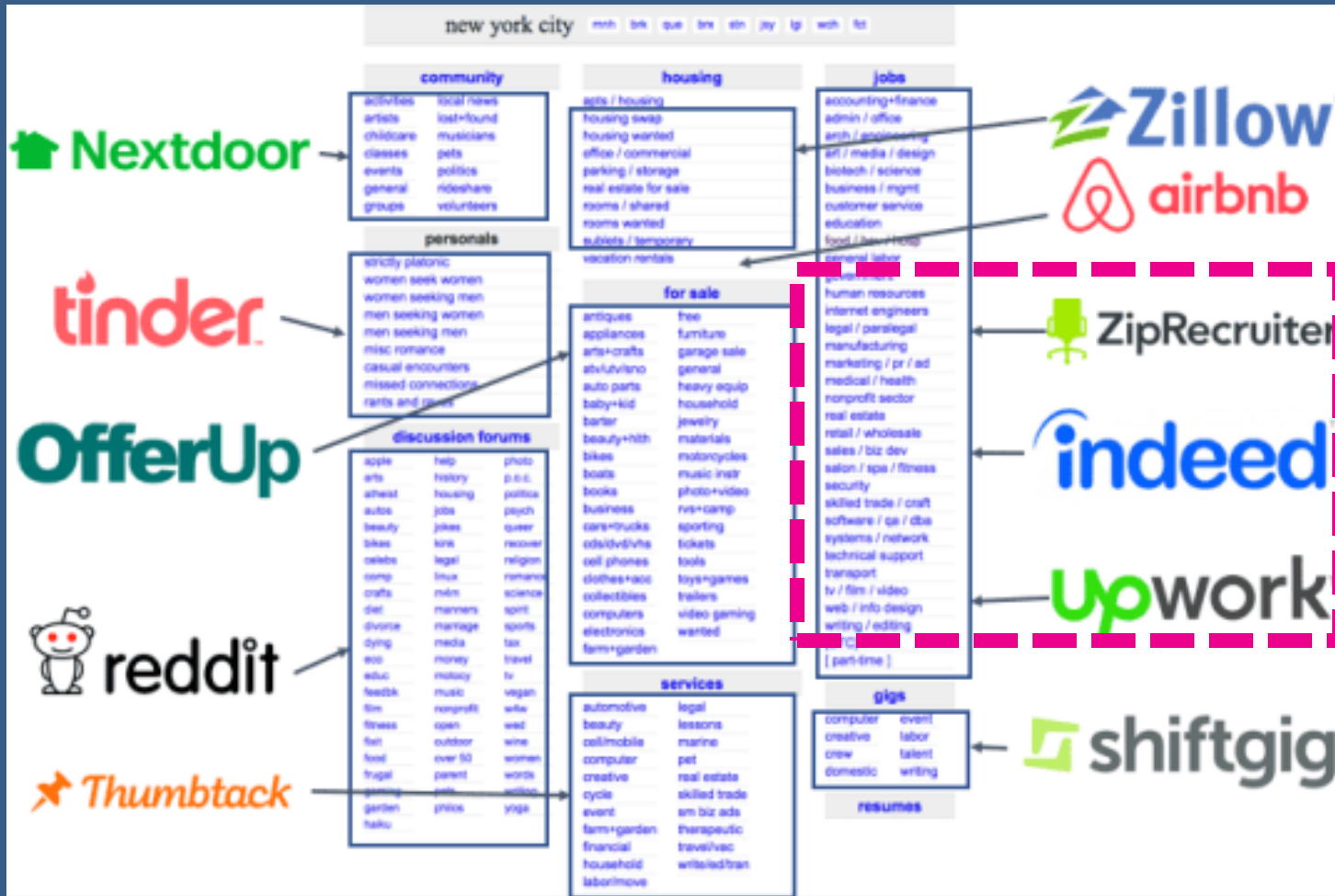


PAIN POINT

JOB POSTINGS

NUMBER OF JOBS
POSTED IS NEAR AN
ALL TIME LOW

COMPETITION



PROBLEM STATEMENT

PROBLEMS



Jobs not classified correctly

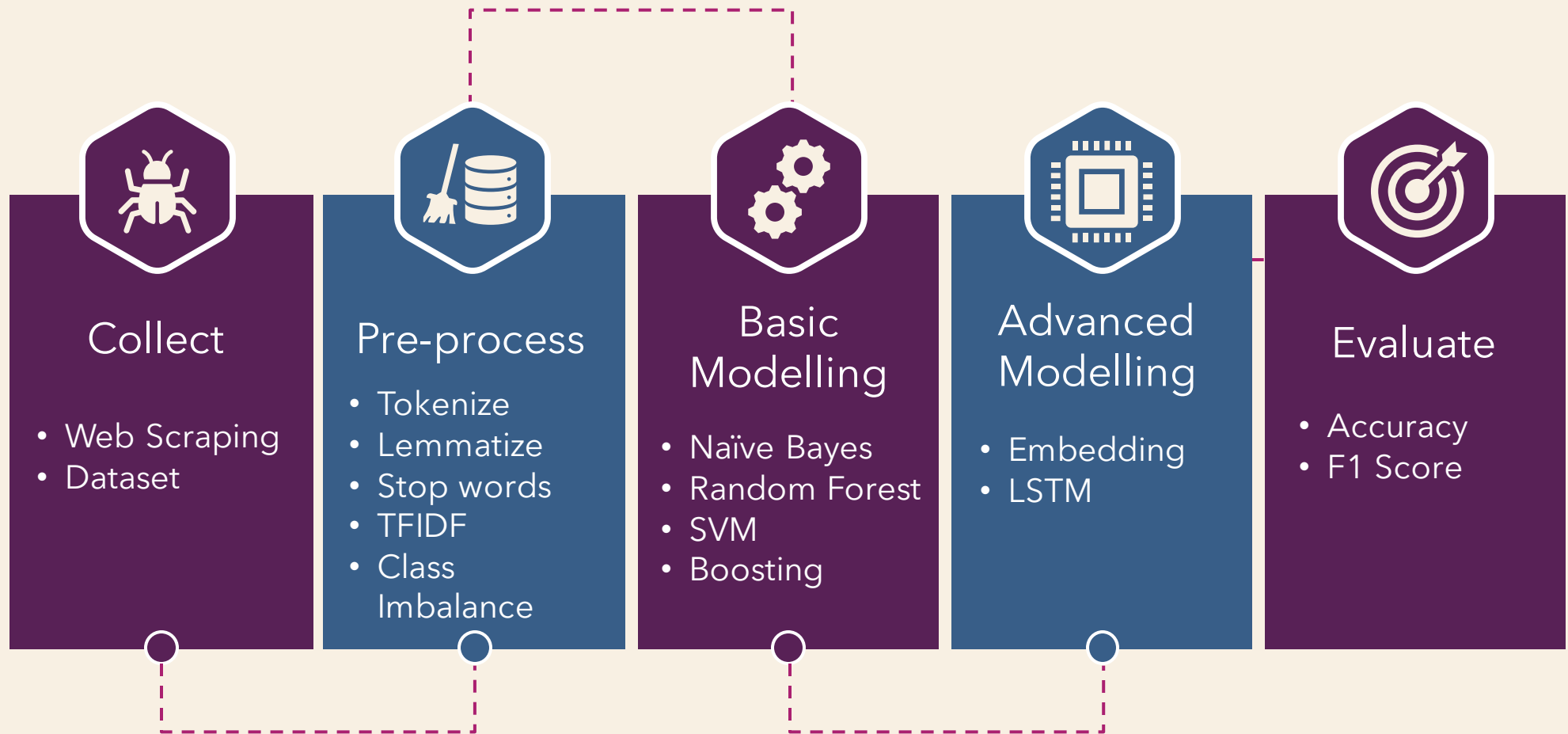


Spam posts reduces brand integrity

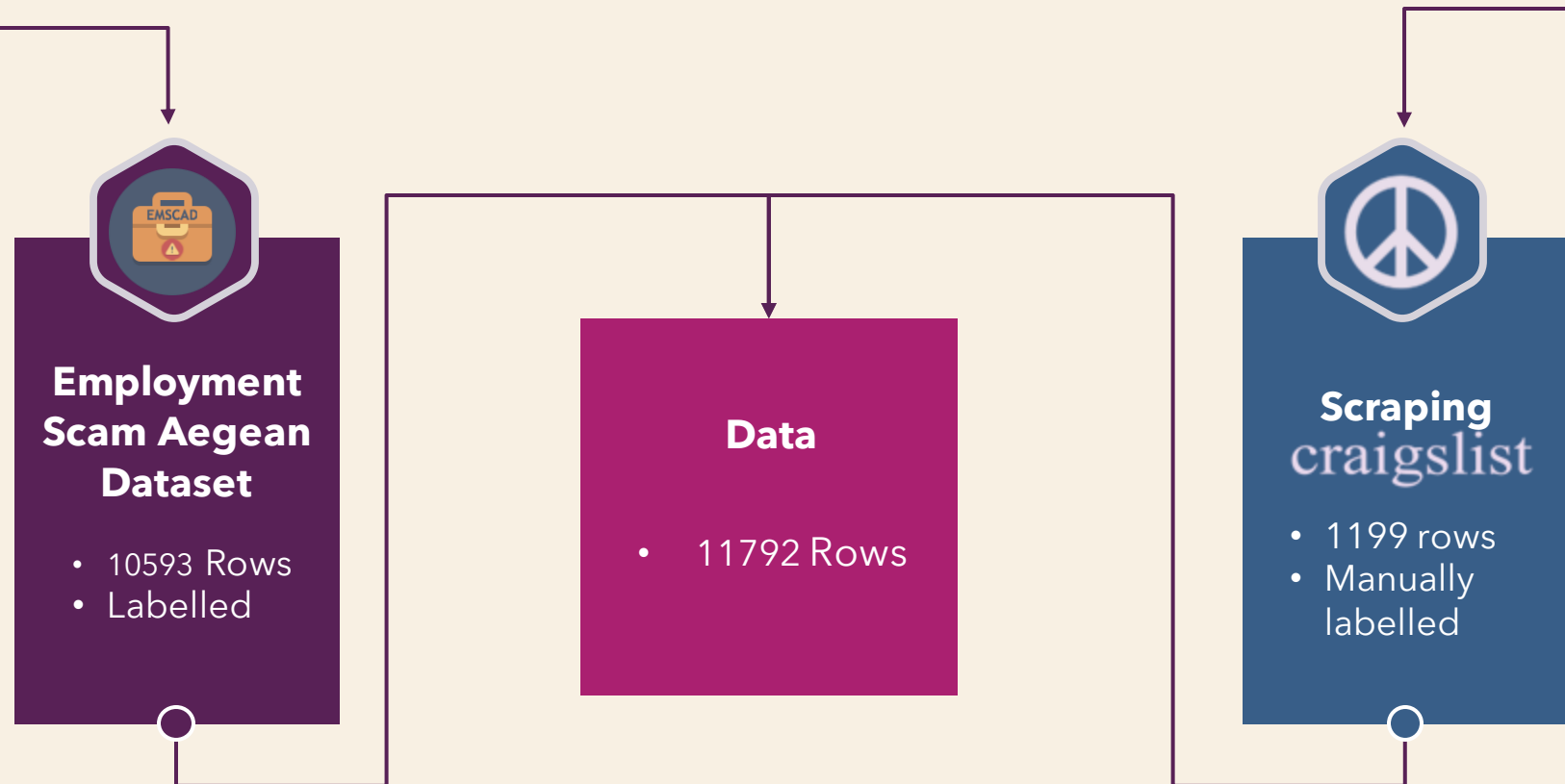
SOLUTION

Build a model to correctly categorize jobs and then go on to detect if the job is fake or not

PROCESS FLOW



DATA COLLECTION



DATA PREPARATION



STEMMING



LEMMATIZATION



STOP WORDS



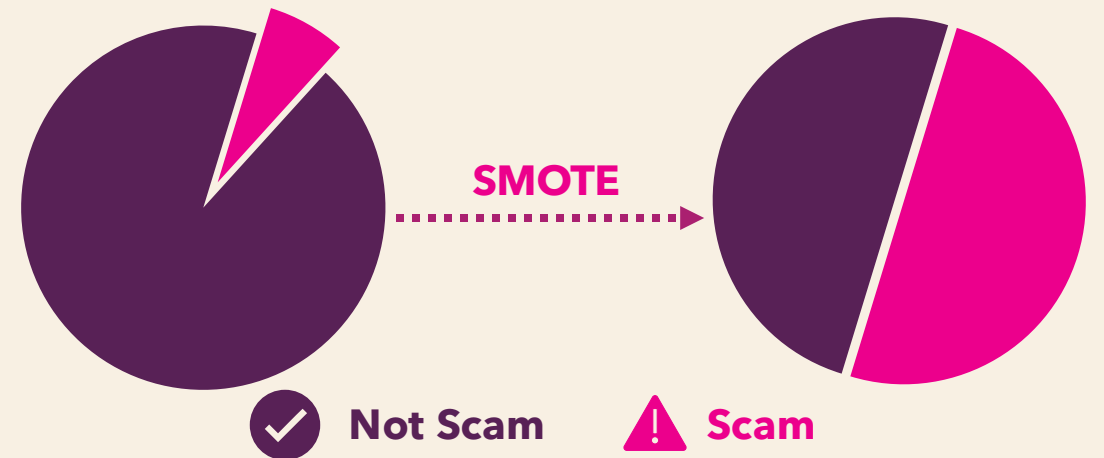
TF-IDF



EMBEDDING



CLASS IMBALANCE



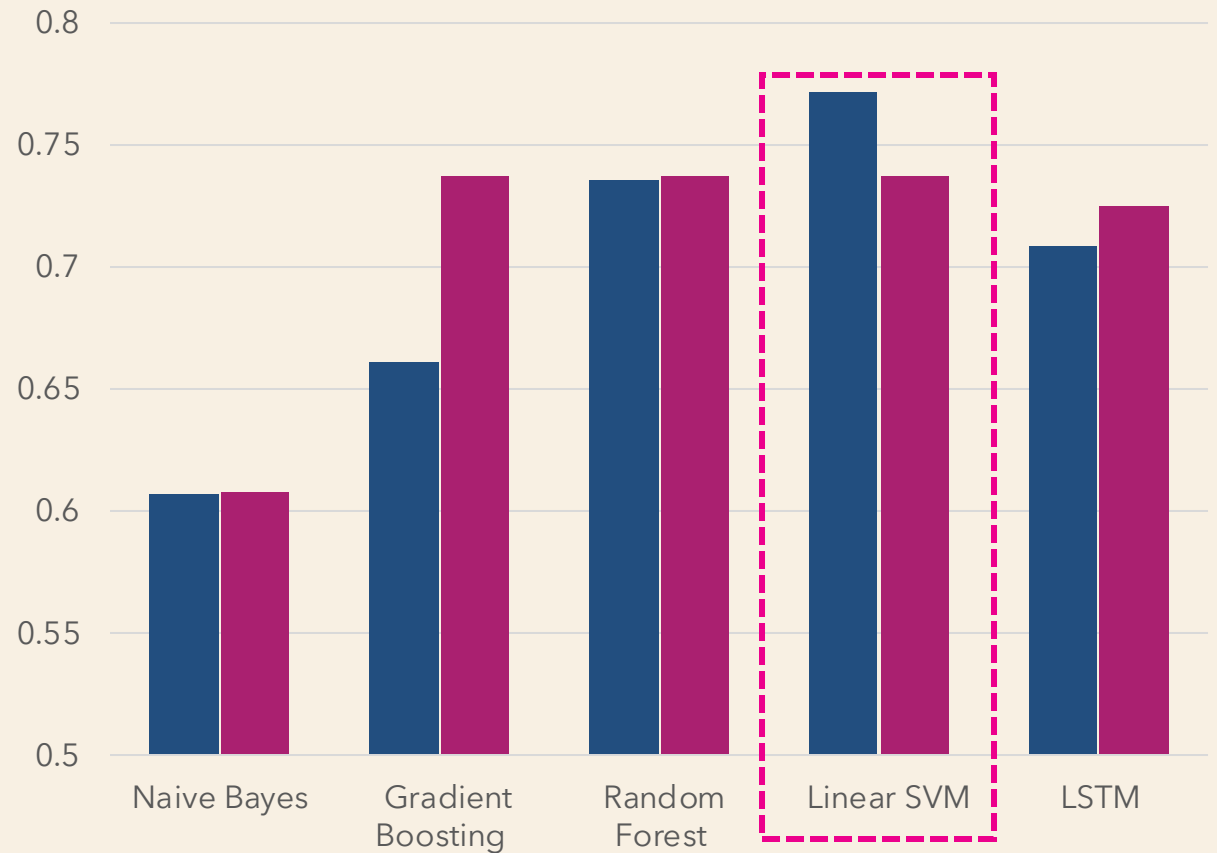
MODELS FOR JOB CATEGORIZATION



Linear SVM with TF-IDF
performed the best

Accuracy: 73%

F1 Score: 0.77



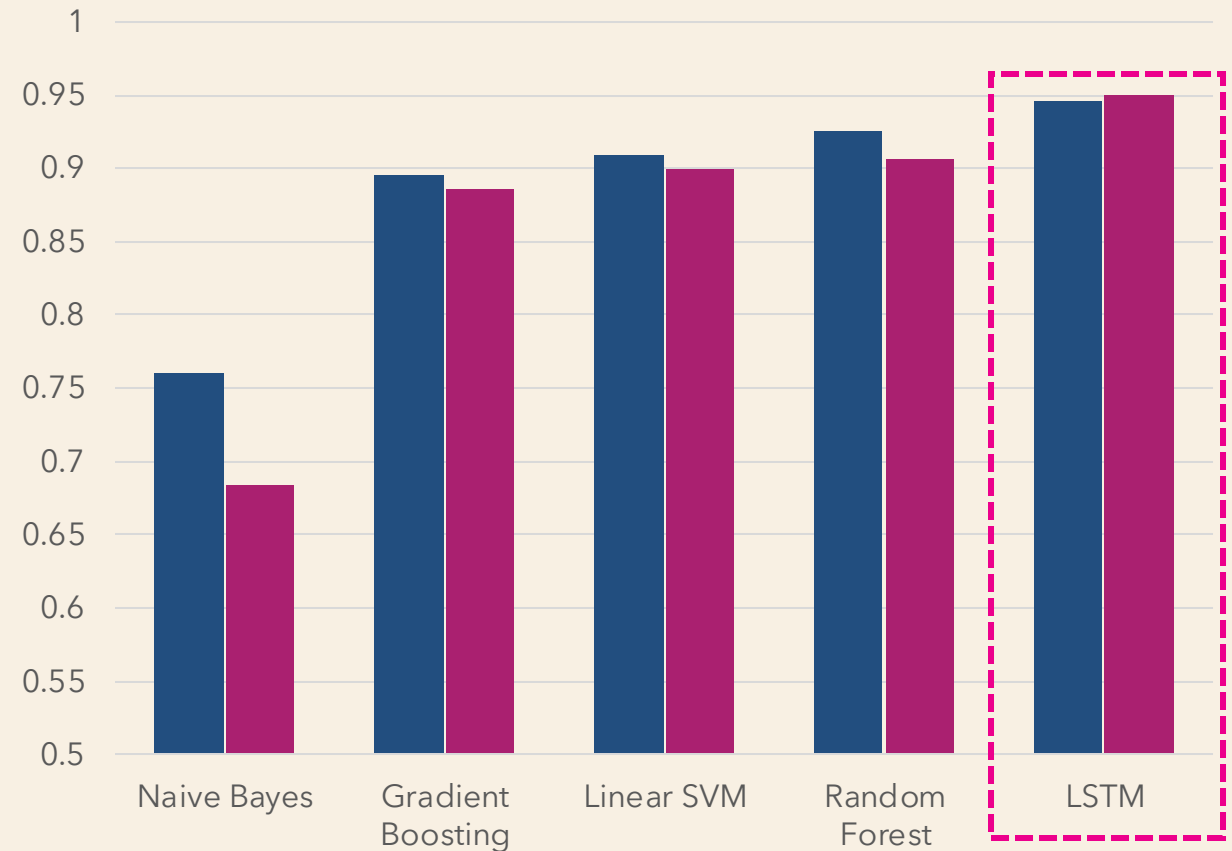
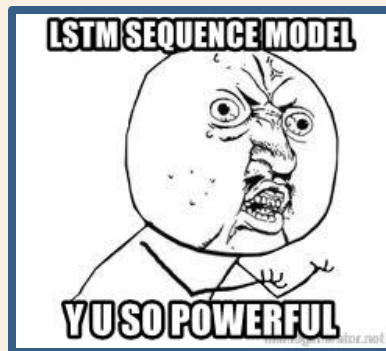
MODELS FOR JOB SCAM DETECTION



LSTM with Word Embedding performed the best

Accuracy: 95%

F1 Score: 0.946



INSIGHTS

SCAM JOBS

Worsens the user-experience which affects all stakeholders. On an average 7 in every 100 jobs posted are scam.

JOB CATEGORIZATION

About 3 jobs in every 100 jobs are wrongly classified. Huge opportunity cost.

ONE SIZE DOES NOT FIT ALL

Model selection should be based on the task, the available data and testing

RESAMPLING

Over-sampling techniques like SMOTE improve model performance by a significant margin

RECOMMENDATIONS

ELIMINATE SPAM JOBS

Use our model to identify spammers and initiate strict action against such accounts. Directly impacts **Brand Image**

BETTER CATEGORIZE JOBS

The opportunity cost of misclassified jobs is high. This impacts **retention** of customers

ADD MORE CHECKS

We see in our analysis that job posts with a company logo are less likely to be a spam

SHIFT FOCUS ON NY. CHARGE \$75 FOR JOBS POSTS

Features of jobs posted in NY are very similar to jobs in SF. Assuming the jobs posted to remain constant, this increases revenue from jobs by **13.65%**

HIRE US

We'll tell you why

FUTURE SCOPE

(IF YOU CHOOSE US)

PRICING STRATEGY

Use data with more variables and advanced statistical models to get the optimum price

CUSTOMER EXPERIENCE

We will understand more variables that have a big impact on the validity of the posting. It's all about the little things.



THANK YOU