University of Texas at Arlington
MavMatrix

2019 Spring Honors Capstone Projects | Honors College

5-1-2019

BETTER ADVERTISING USING SMARTER SCREENING

Divyanshu Sharma

Follow this and additional works at: https://mavmatrix.uta.edu/honors_spring2019

Recommended Citation

Sharma, Divyanshu, "BETTER ADVERTISING USING SMARTER SCREENING" (2019). *2019 Spring Honors Capstone Projects*. 3.
https://mavmatrix.uta.edu/honors_spring2019/3

This Honors Thesis is brought to you for free and open access by the Honors College at MavMatrix. It has been accepted for inclusion in 2019 Spring Honors Capstone Projects by an authorized administrator of MavMatrix. For more information, please contact leah.mccurdy@uta.edu, erica.rousseau@uta.edu, vanessa.garrett@uta.edu.

Copyright © by Divyanshu Sharma 2019

All Rights Reserved

BETTER ADVERTISING USING
SMARTER SCREENING

by

DIVYANSHU SHARMA

Presented to the Faculty of the Honors College of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

HONORS BACHELOR OF SCIENCE IN COMPUTER SCIENCE

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2019

ACKNOWLEDGMENTS

I would like to thank my mentor, Dr. Christopher Conly, for supporting me throughout the planning and development phase of the project. Next, I would like to thank each member of my team; Eric Sauber, Micah Clarke, Mehul Jain, and Sakshi Nag, for their commitment and effort from the start to the end.

I would also like to thank advisors of the Honors College for helping me navigate and understand all the different processes.

May 03, 2019

ABSTRACT

BETTER ADVERTISING USING

SMARTER SCREENING

Divyanshu Sharma, BS Computer Science

The University of Texas at Arlington, 2019

Faculty Mentor: Christopher Conly

Advertisers are moving away from traditional advertising media and focusing on Social Media platforms with the help of Social Media Influencers. These influencers create “sponsored” content that seems out of place. This reduces the users’ quality of experience and appears forced.

BASS ensures that an advertisement would blend naturally and smoothly into an influencer's social media by analyzing the similarities between the advertiser and the nature of the influencer's past content. BASS involves using tools, such as Computer Vision and Data Mining, to generalize the trends among brands, themes, and objects that recur in the influencer's past social media posts. Once these trends are obtained, a text vectorization and matching algorithm ranks different influencers in order of natural fits for a given advertiser.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 Purpose

The purpose of BASS is to find and implement a way to match advertisers with credible influencers using an array of technologies acting on the influencers' Instagram profiles. This makes advertising a more natural and smooth experience.

1.2 Motivation

On social media platforms, there are times when influencers advertise content that simply does not flow well or match with what their profile or theme is truly about. For example, an influencer who built her reputation on cooking videos advertising car parts from an auto parts store doesn't really make much sense to their followers. Online advertising, and advertiser-influencer matching, is an area that needs automation to truly exploit the huge number of followers that these influencers amass.

1.3 Scope

For the sake of this project, and to limit the complexity, team BASS decided to limit ourselves to Instagram - a popular photo and video sharing platform plagued with new influencers. BASS provides top matching influencers for any advertiser given a set of keywords or search query. BASS also provides a series of filters and options to further reduce the size of results.

1.4 Approach

A good way to ensure that an ad smoothly blends into an influencer's social media is to analyze the nature of influencer's past content. This involves using tools, such as Computer Vision and Data Mining, to generalize the trends among brands, themes, and objects that recur in the influencer's past social media posts. We obtain social media posts by scraping the influencer's public profiles and through the Instagram API. Once we obtain these trends, a text matching algorithm helps rank different influencer in order of natural fits for a given advertiser.

Team BASS used Agile Development methodology, which relies on delivering results in every sprint (2 weeks) with a Scrum Master keeping track of development progress. The results of each sprint are integrated into the master branch after integration testing.

1.5 Intended Audience

Intended audience includes companies and individuals looking to increase their online outreach through advertising on various social media outlets.

CHAPTER 2

REQUIREMENTS

Team BASS brainstormed requirements along with our mentor Dr. Christopher Conly during meetings and the throughout the first semester of Senior Design.

2.1 Customer Requirements

The mentor and team members of team BASS act together as customers and formulate requirements that are essential for the project's success.

2.1.1 Web Scraping

2.1.1.1 Description

The system must be able to extract information from a given web address with minimal errors. The system should be able to scrape the web address, extract information, and channel that information to the system to procure the best recommendations possible.

2.1.1.2 Source

The source of this requirement is the group discussion and base requirements needed to run the system efficiently with fewest errors.

2.1.1.3 Constraints

The constraints can be restricted access to a web address or profiles, which are locked by default. Also, the permission of the social media influencer to access their profiles is crucial to the project.

2.1.1.4 Standards

The system should be able to scrape information from HTML 5 web pages.

2.1.1.5 Priority

Critical

2.1.2 Download Photos and Tags

2.1.2.1 Description

The system must be able to download the photos and tags from the Instagram profile. It should be able to recognize the photos and classify them. It should also be able to scan the images and recognize common objects and brands in them.

2.1.2.2 Source

The source behind this requirement is the idea behind this project and the description of what our software will accomplish.

2.1.2.3 Constraints

We should be able to gather a large database with brands and their logos. Also, we should have product descriptions, sizing, etc., to identify them in the photos and posts on social media.

2.1.2.4 Priority

High

2.1.3 Generate Tags from Brands and Objects

2.1.3.1 Description

The system must be able to categorize entities in downloaded images into identifiers such as logos and brands, which will help in provide better recommendations.

2.1.3.2 Source

The source behind this requirement is the idea behind this project and the description of what our software will accomplish.

2.1.3.3 Constraints

The tags should be in text format.

2.1.3.4 Priority

High

2.2 Performance Requirements

Computer Vision plays a crucial role in the recommendation process because it is the main entity recognizing and storing the photographs and identifying the specific brands. It is important for this process to be as fast and as efficient as possible.

2.2.1 Computer Vision

2.2.1.1 Description

Operations include using Computer Vision to recognize objects and brand's logos in those photos. The brands and products in the photographs will then be analyzed with respect to their relative popularity to other influencer's uploads and the influencer's own posts to optimize and increase likes and popularity.

2.2.1.2 Source

Team BASS and our mentor.

2.2.1.3 Constraints

An efficient computer vision process is needed to minimize errors when scraping images from user's social media pages. Speed and accuracy of recognition results will be a major factor in the success of BASS. Other factors include pages with security settings set to private.

2.2.1.4 Standards

Accuracy will be determined manually by team when reviewing the data that the application has scraped.

2.2.1.5 Priority

High

2.3 Laboratory and Robotic Requirements

2.3.1 Laboratory Usage Requirements

2.3.1.1 Description

Any fabrication equipment provided used in the development of the project shall be used in accordance with OSHA standard LOTO procedures. Locks and tags are installed on all equipment items that present use hazards, and ONLY the course instructor or designated teaching assistants may remove a lock. All locks will be immediately replaced once the equipment is no longer in use.

2.3.1.2 Source

CSE Senior Design laboratory policy

2.3.1.3 Constraints

Equipment usage, due to lock removal policies, will be limited to availability of the course instructor and designed teaching assistants.

2.3.1.4 Standards

Occupational Safety and Health Standards 1910.147 - The control of hazardous energy (lockout/tagout).

2.3.1.5 Priority

Critical

2.3.2 RIA Robotic Manipulator Safety Standards

2.3.2.1 Description

Robotic manipulators, if used, will either housed in a compliant lockout cell with all required safety interlocks, or certified as a "collaborative" unit from the manufacturer.

2.3.2.2 Source

CSE Senior Design laboratory policy

2.3.2.3 Constraints

Collaborative robotic manipulators will be preferred over non-collaborative units in order to minimize potential hazards. Sourcing and use of any required safety interlock mechanisms will be the responsibility of the engineering team.

2.3.2.4 Standards

ANSI/RIA R15.06-2012 American National Standard for Industrial Robots and Robot Systems, RIA TR15.606-2016 Collaborative Robots

2.3.2.5 Priority

Critical

CHAPTER 3

ARCHITECTURE AND DESIGN

The figure below provides a high-level overview of the BASS system architecture. It is comprised of five key components that interact with one another by sharing, storing, and retrieving information. System starts with the front-end web application that includes the user interface to provide an interactive interface to the service BASS offers. The web application contains various features and functionality described below that connect to different modules of our service. There are three software modules that are built around a centralized database. Each piece of software has entirely independent functions and capability, but they all play a role in determining the end result of our service.

The flow of the system begins with an advertiser signing up for the service via web application by providing required information used by the system that is stored in the database. After the advertiser agrees to move forward with the service, the retrieval module scrapes information from the web that is then saved to the database. The control is handed off to the recognition module, which is in charge of learning what different objects/brands look like and recognizing/parsing information. This information is sent back to the database to be utilized by the matching module. This matching module calculates a score related to the advertiser’s product/service. This score is sent back to the front-end user interface where the advertiser can view the search results.

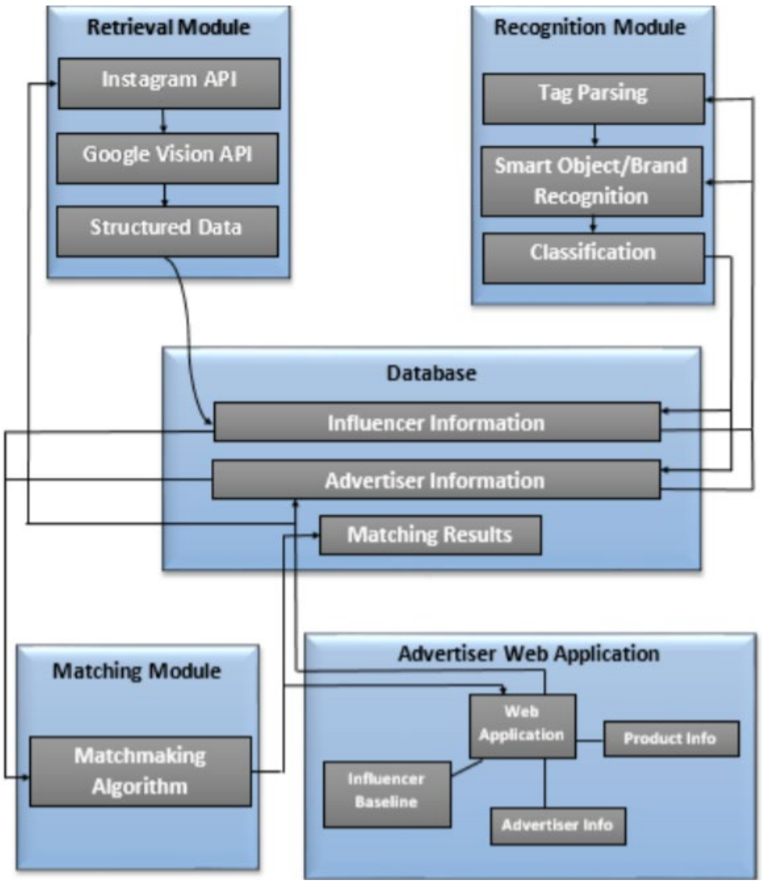

Figure 3.1: Overall structure of BASS

3.1 Database

3.1.1 Database Layer Subsystem

The NoSQL MongoDB database is responsible for storing advertiser, influencer, classification, and matchmaking result information from the web application and three modules of our system. Images were not directly stored. Instead, public URLs of the images were stored to save storage space.

The database is accessed by all software modules and the web application, so access control of user accounts, tables, stored procedures, and safe programming practices can be implemented to safeguard against outside threats. Clean-up jobs and maintenance tasks are set-up in server agent to clean up influencer's "trash" photos that do not correlate to the overall theme of the profile. In addition, if an influencer deletes their profile then all data relating to the influencer is deleted to free up space.

3.1.1.1 Layer Hardware

The database is hosted on MongoDB Atlas, which is a fully automated cloud service that is deployable on the Google Cloud Platform.

3.1.1.2 Layer Operating System

MongoDB is a cross-platform document-oriented database. Operating system of the cloud service is unknown and does not affect the project's implementation in any manner.

3.1.1.3 Layer Software Dependencies

It requires djongo, which is a connector that allows you to use Python's Django with MongoDB without changing the Django ORM. The database version is v4.0.6.

3.2 Retrieval Module Subsystem

The Retrieval Module is tasked with filling out our influencer portfolio with images from the influencers' profiles, profile statistics, and general account information to be exported to the database. The retrieval module searches for public Instagram accounts that satisfy a certain threshold of followers. After analyzing accounts that meet the criteria, the URL for those accounts are stored in the database. The recognition module then scrapes those influencers' images, tags of images, account information, and statistics via the URL stored in the database.

3.2.1 Instagram API

With an Instagram user's permission, our developers can access the contents such as posts, the associated comments, and tags from the Instagram user's profile.

3.2.2 Google Vision API

This subsystem coincides with the Recognition subsystem. More detailed information can be found there.

3.2.3 Structured Data

This subsystem structures the data obtained from Google Vision API to ready it for storage in the database.

<u>3.3 Matching Module Subsystem</u>

The Matching Module imports Instagram information of influencers that an advertiser may want followed by classifications of each influencer using the recognition module. This information is extracted from the database. The data from both the advertiser and influencer is run through the matching algorithm that cherry-picks potential candidates that meet the mold provided by the advertiser. These results are sent to the database and the web application where they are displayed.

3.3.1 Layer Hardware

We use an Nvidia graphic card to run and train our model and use it to match the influencer with the favorable brand. The hardware plays a key role in the speed and the accuracy of the model. This is essential because it captures the essence of the project. Additionally, we use Google Cloud Vision services to enhance brand and logo recognition as well.

3.3.2 Layer Operating System

It can run on any OS, which satisfies the requirements to run Python 3.6.

3.3.3 Layer Software Dependencies

Python 3.6, scikit learn, NumPy.

3.3.4 Subsystem - Match Recommender

It recommends the specific brands, which the influencer should endorse so that the joint venture is profitable to both the advertiser and the brand.

3.3.5 Subsystem Hardware

An adequate graphic card and PC capability to support the algorithm.

3.3.6 Subsystem Operating System

OS Independent.

3.3.7 Subsystem Software Dependencies

Python 3.6 and its libraries.

3.4 Advertiser Web Module

This module displays preferences and results that can be specified by the advertiser including tags, Instagram descriptions, and match ratings returned from the matching module. This web application is an interactive user interface for the advertiser to view and upload information. This is the main point of interaction between the advertiser and BASS.

3.4.1 Layer Hardware

N/A

3.4.2 Layer Operating System

It can run on any OS which satisfies the requirements to run Python 3.6

3.4.3 Layer Software Dependencies

Python 3.6 and its libraries.

CHAPTER 4

RESULTS AND CONCLUSIONS

BASS is a complete product that can be easily deployed on a website and sold as service. Given a suitable string of keywords and additional requirements, such as minimum number of followers, an advertiser can see a set of top three matching Instagram profiles along with the matching score as a percentage. The user interface is interactive, easy to navigate, and minimalistic.

BASS relied heavily on open source resources and tools that helped reduce the cost of implementation. Due to the team members' student status, team BASS secured free hosting services and did not use any of $800 budget that the College of Engineering provides each Senior Design team.

BASS dealt with significant lack of support and drastic redesign of the Instagram API following massive privacy scandals and password thefts at Facebook. Despite the challenges, team BASS recovered and implemented the project before the deadline.

BASS combined several areas of expertise of all the team members. Additionally, we also gained knowledge about technologies such as Computer Vision and Data Mining.

BASS successfully automated a huge part of influencer-advertiser matching while accounting for factors such as the right fit and engagement index of the influencer's followers. BASS incurred no infrastructure cost due to several discount and free options for new developers and computer science students. BASS successfully met and exceeded all the requirements set in the initial stages.

REFERENCES

Berman S.J., Battino B., Shipnuck L., Neus A. (2009) The End of Advertising as We Know It. In: Gerbarg D. (eds) Television Goes Digital. *The Economics of Information*, Communication and Entertainment (The Impacts of Digital Technology in the 21st Century), vol 01. Springer, New York, NY

Bullock, Lilach. How to Evaluate and Partner with Social Media Influencers. *Social Media Examiner*, 31 July 2018.

Freberg, Karen et al. “Who Are the Social Media Influencers? A Study of Public Perceptions of Personality.” *Sciencedirect*, Academic Press, 4 Dec. 2010.

How to Reach Out to Influencers So That They Can’t Say No. *The Daily Egg*, 8 June 2018.

Newberry, Christina. Influencer Marketing on Social Media: Everything You Need to Know. *Hootsuite Social Media Management*, 10 July 2018.

BIOGRAPHICAL INFORMATION

Divyanshu Sharma is a student at the University of Texas at Arlington. He enrolled in the Bachelor of Science in Computer Science in Fall 2015 and joined the Honors College in his sophomore year. He participated in major research projects with the Center for High Energy Physics, Department of Physics and the College of Engineering while pursuing his degree. He plans to work and gain professional and life experience before pursuing an MBA.