UNIVERSITY OF EDINBURGH
Business School

2020-21

[CMSE11427 & Web and Social Network Analytics]

[Individual Project]

[B167586]

Part 1:

Recommender systems are used to predict the 'rating' a user would give an item they have not already rated. Recommender systems have two forms collaborative filtering-based (unsupervised) and content-based (supervised). This report focuses on collaborative filtering-based systems which aim to provide 'rankings' for a user's preferences based on the user's past behavior and the behavior's of other users. The behavior's of other users who are considered 'similar' to the user of interest have a higher weight in determining what future items will be suggested.

To explain collaborative based filtering's usefulness in the matching process for candidates and jobs assume a *nxn* matrix where the x-axis contains job listings and the y-axis contains candidate profiles. The matrix is populated by 1's (candidate has applied for job) or 0's (candidate has not applied for job). The similarity measures (Jaccard or cosine) would then be used to suggest which jobs a candidate should consider applying for based on how similar their application history is in comparison to the other candidates in the data base. If the matrix was populated by 1's or 0's depending on if a candidate had received an interview this could be used by employers to suggest who may be good candidates for a company to offer an interview.

The benefits of collaborative based-filtering is that new jobs, they otherwise may not have seen, will be suggested to candidates. In context of this report, a con is that if there is a candidate that has not received any interview offers or a job listing that has received no applications, these observations cannot be evaluated in the model. The model will not suggest the job to other candidates or the candidate to employers because they are not 'embedded' in the model.

This shows how recommender systems can be echo chambers where only certain people are recommended because they are similar to people who are traditionally seen. Recommender systems are not aware of systematic deprivation that has for decades left many potentially great workers at an unfair advantage. Generational wealth (or lack of it) due to race or geographical location and gender are some of the largest inhibitors for young professionals. This will reduce the diversity (racial, gender and socioeconomics background) of candidates for a job thus reducing much needed heterogeneity in the work force. Skilled HR employees should purposefully seek out diversity, overriding the recommender systems when needed.

Part 2:

It appears an influencer is someone who creates their network, fanbase and livelihood through a social media platform while a celebrity creates their network, fanbase and livelihood in a separate sector but then joins the social media platform. Nevertheless, the same metrics: *degree*, *betweenness* and *authority* are used to quantify an online profile. In regards to a influencer's social media account, degree centrality measures the direct connections an influencer has, either directed (Instagram/Twitter followers) or undirected (Facebook friends). Betweenness centrality indicates what influencers act as a link between different sub-groups or social networks on a platform. Authority measures the quality of content for an influencer's account.

These metrics are best when viewed in a wholistic process, none is the single best evaluation tool. Degree is a good indicator of the immediate viewership an influencer could have. High betweenness could lead to more individuals coming across an influencer's account in an 'organic' fashion. As a potential customer scrolls/clicks through a social media platform there is a better chance they will come across an influencer with higher betweenness than an influencer with lower betweenness. An influencer with a high authority score ensures that when a company's product is sponsored on their account it will be viewed as a quality product. All three metrics share one weakness. None track engagement: likes, comments, tags, etc… Fan engagement could indicate followers are more likely to interact and trust the products an influencer endorses. An influencer with millions of followers who scroll by their posts without paying close attention to the content would want to be avoided if possible.

To avoid the echo chamber finding influencers with potential for diverse branding is key. For example, an influencer who posts daily history factoids but in their spear time focuses on fitness as a means of mental health preservation. Once a week they post a fitness update in addition to their normal history posts. While the diversification is subtle it could help expand their audience base in terms of degree and betweenness. Adding different wrinkles to an influencer's profile can make them more appealing to the masses.

<u>Part 3:</u>

Glossary:

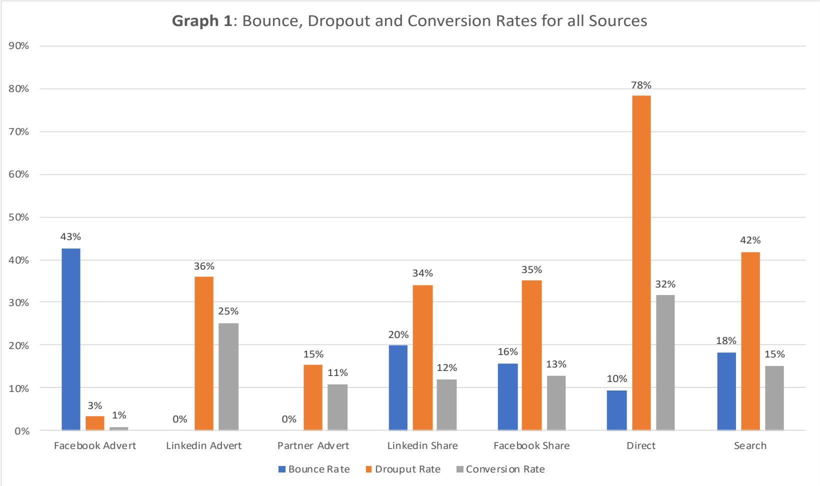
Bounce Rate: Percentage of clickstreams with length 1

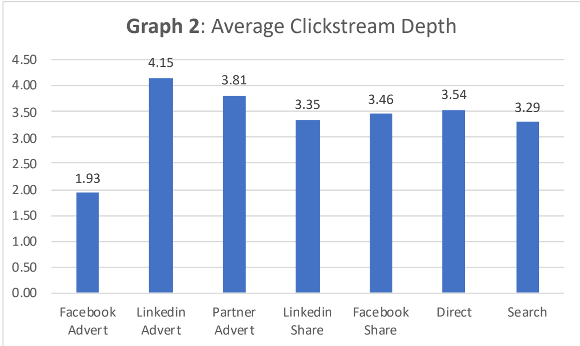
Dropout Rate: Percentage of purchases completed once started

Conversion Rate: Percentage of purchases completed out of all relevant clickstreams

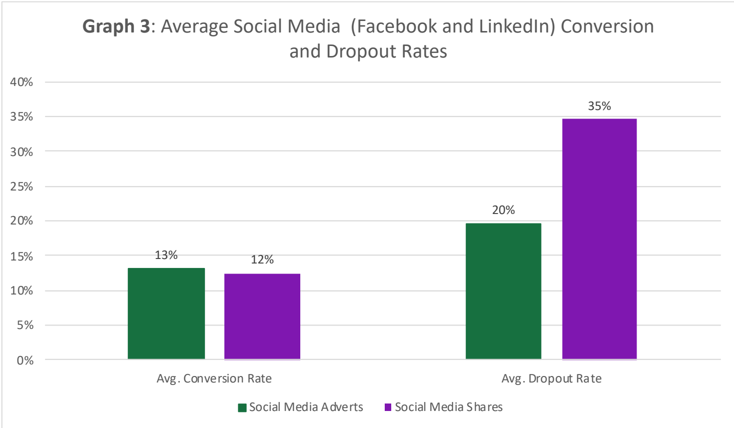
Average Clickstream Depth: Average number of pages visited per relevant clickstreams

Based on bounce rate, dropout rate, conversion rate and average clickstream depth LinkedIn had the best advertisement campaign. See **Graph 1** and **Graph 2** below. Conversion rate indicates that LinkedIn advertisements had the most successful purchases per use. The dropout rate indicates once people begin buying product, if there source is a LinkedIn ad, they are more likely to complete the purchase than if the source were other advertisements. LinkedIn also has the highest average clickstream depth; customers engage with the website longer. LinkedIn is tied with the lowest bounce rate; customers never immediately leave after visiting one page.

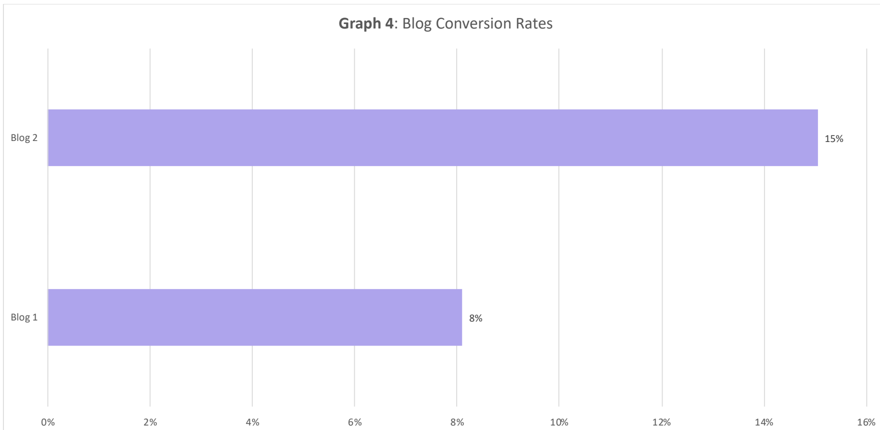




A comparison of social media (SM) advertisements vs. SM shares was conducted. Conversion rate performance SM advertisements (Facebook and LinkedIn) are better than SM shares. The average SM advert conversion rate is 13% compared to SM shares 12% rate. However, the average dropout rate is 20% for adverts and 35% for shares. Thus, SM adverts create more purchases than shares per use but SM shares have more individuals completing purchases once started. See **Graph 3** .



Additional content like blogs can improve website experience, thus, each blog was studied. *Blog 2* converts better than *Blog 1*, 15% vs. 8%, respectively. **See Graph 4.** Therefore, blogs similar to *Blog 2* should be used. This metric was calculated by finding the total number of times *Blog 1* or *Blog 2* were mentioned in the clickstream before a purchase success occurred. If neither blog or both blogs were visited before a purchase success the clickstream's information was not included in the calculation.



Customer behavior from each platform after the purchase process began was investigated. The most common final page visit for each platform, when a purchase was started but not completed, was found. *Android*, *mac*, *unknown* and *windows* shared their most common end page, 'purchase start.' Their second most common final page visit was, 'purchase enter address.' For *ios* the most common end page is 'purchase enter address' and the second most common is 'purchase start.' **See Graph 5.** This indicates that aside from possible accidental purchase starts or 'cold feet' there are issues when customers need to enter their address.