Social Scoring – Good or Bad?

Adeela Huma
ahuma@vt.edu

ABSTRACT

With traditional credit scoring data such as FICO , there are between 45 and 64 million [2] Americans without a credit score and this includes mostly recent college graduates and immigrants. They are often denied credit because there is not enough data to make a decision using the FICO system. The current economy is in the need of an alternate scoring system to score these 'unscoreable' consumers so that Banks can give more loans and consumers can fulfill their needs. To fill this gap Big data comes to rescue where companies like Friendly score [4] or Lenddo [3] use social media data to provide the social credit score of a person. Based on this social score loan applications are approved or denied. This paper analyses the ethical consequences of using Friendly score API.

KEYWORDS

FICO score, Social Score, PYTFund, Friendly score, GDPR, Bid Data, Data Breach, PII

1 INTRODUCTION

FICO is the gold standard for calculating credit scores. Loan applications, credit card inquiries etc. all essentially depend on one thing: a credit score. Traditionally, institutions have asserted that an individual's financial history (primarily payment history, current debt profile, and length of credit history) informs future behavior and therefore should dictate the decision to offer credit to potential borrowers and at what rate. This framework makes sense when the necessary information is available, but what about when it is not? That's why low-income people, immigrants, students have hard time to get financial applications scored.

With traditional credit scoring data such as FICO , there are between 45 and 64 million [2] Americans without a credit score and this includes mostly recent college grads and immigrants. They are often denied credit because there is not enough data to make a decision using the FICO system. Not surprisingly, about 40% of these "unscoreable" people are homeowners, and many hold professional-level jobs or are retired. That means if you are a student and have no cosigner chances of getting loan are very slim. That also means that a significant portion of these "unscoreable" Americans are potential customers for lenders offering loans and lines of credit.

Banks make money when they collect interest on the loans they make and credit cards they issue – so having more people to sell their services to increases their profits. To score these "unscoreable" consumers, they need other sources of data and here comes Big Data for rescue.

Bid Data provides alternate data options compared to FICO score to identify potential consumers who otherwise would be rejected for loan. Some lenders are already using cable bill history, rent history, cell phone bill history, and payday loan history to qualify consumers with "thin" credit files. And once the bank has identified you as a potential borrower, they can snoop in your social media profiles for more information to judge your creditworthiness. It's all perfectly legal because what you post on social media is public information. All your posts, tweets, and selfies are free for anyone to see.

Social scoring works a bit like credit score but instead of being based on good or bad financial history, scoring is done based on social media information of user. Its predicting individual's character or willingness to pay based on social network data. There are a few companies such as Lenddo [3] and Friendly Score [4] who use social media data to generate social score that is used by lenders to give loan.

This paper analysis the ethical implications of using social media data to give loans. In this paper, Friendly score API is analyzed to identify if social media scoring has the potential to be discriminatory towards protected categories, majors etc. Is social media scoring actually a FICO score in disguise only the data points are changed? Does people who have strong social profile are able to get loans compared to people who are trustworthy people but has less social presence because they are not tech savvy. The analyses is done only on Friendly score data made available by PYTFund [5] organization.

1.1 PYTFund

PYT (Pay your tuition) Fund [5] organization finances students who are close to graduation and need funding for less than $5000. The applicant sends application to PYT and it forwards all of them to lender (in this case GS2). PYT Fund generates a "credibility or risk score" for student based on which it can convince the lender to provide loan to student or provide loan from its money pool as long as score is less than national default rate. Two main scoring components are: a) **Crowd funded** money from applicant's

social circle – the more money applicant gathers from social circle the higher their score will be. b) **Student's social score** - Social network based score.
To acquire student's social score PYT is using "friendly score API" [4]. The higher score reflects greater credibility. Since this study is done in collaboration with PYTFund, friendly score API is analyzed to understand if its usage is ethical or not.

1.2 Friendly Score API

Friendly Score API [4] uses social network information to generate scores and provides information to lenders via API. Social networks used are:

- Google
- Facebook
- Instagram
- LinkedIn
- PayPal
- Twitter

Friendly Score API has two types of responses:

- aggregated score for each social platform
- details of students on each platform

Friendly score's 'score range' is from 0 to 1000 points. The higher score reflects greater credibility. Lenders use Friendly score API to get the social score of students. Students who want to provide social score to lenders login into lenders portal (PYTFund) or directly to Friendly score website and provide information about their social profiles. Friendly score asks student the permission to access their social media data. The student is free to provide information about one or all of its social media profiles. At the end Friendly score will 'score' student based on given information (social media profiles). Later lenders can access information about student's score via friendly score API's token. Friendly score's 'score' is not a fixed score, it changes depending on information in student's profiles.

1.3 Friendly Score API Analysis

To analyze which factors contribute to higher social score, this study looks answers for some open ended questions such as:

- Does protected categories such as gender, ethnicity play any role in score calculation?
- How major affects friendly score ?
- How GPA affects friendly score?
- How number of friends affect friendly score?
- Does social score varies depending on the number of sources of data i.e. twitter, Facebook etc.?
- Which source (Facebook, twitter etc.) contributes more in social score?
- Does being employed affects friendly score ?

Based on these questions this paper draws conclusion that if it is ethical to use friendly score API to gives loan based on social score.

2 Dataset

The multiple sources of data is used for analysis. One dataset used is from PYTFund. The PYTFund is an organization that provides loan to students based on their social media score from Friendly score API. This dataset has around ~6500 records
Second dataset is from Friendly Score API. As mentioned in previous section lenders can get student's score from friendly score by using their token. By using PYTFund's token of friendly score, scored data is collected for all student's that gave approval to PYTFund to access their social score. This dataset has around ~70 records.
For analysis both datasets are merged that gives around ~26 records after excluding data with missing values etc. This is definitely not a lot of data but this is all was available for analysis.
Another approach considered was to create fake profiles of students with varying attributes such as gender, ethnicity, zip codes, names etc. to understand contributing factors in Friendly score's social score. But this didn't come out to be feasible short term approach because friendly score's social score was almost same in initial stages and it appears that more active and old profiles have good comparable score. Due to these reasons PYTFund and Friendly Score data is used for analysis.
The code used for data munging and analysis can be found at this git hub location [].

3 RESULTS AND DISCUSSION

This section describes the analysis of different attributes (such as gender, ethnicity, major, GPA, number of friends, number of profiles provided , employee etc.) on social score from Friendly score API.

3.1 Protected categories

Protected categories such as race, gender are sensitive data points and their usage in big data has been found discriminatory[6]. PYTFund dataset has these attributes (gender, ethnicity) but has missing values in current data set. On the other hand Friendly score API does not explicitly returns these attributes. Gender does exist in returned payload collected from social networks but is not always present. So there is no data that can be analyzed to infer that protected attributes play any role in score calculation.

3.2 Major

Since PYTFund gives loans to student, does certain Major gets higher score compared to others. The Majors like IT and Medical gives students high prospects of getting jobs. So maybe students who opt for these majors will have higher friendly score.

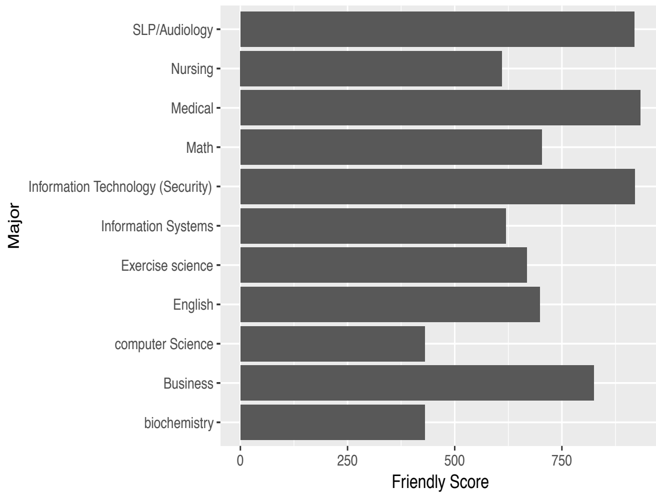

Figure 1: Major vs Friendly score

Another thing one would assume is that Majors under certain disciplines such as Health (Medical, Nursing, Exercise science) or IT (Computer Science, Information Systems, Information Technology) would have similar or comparable scores but figure 1 shows that with even the same discipline score is different that means Major is not the contributing factor in friendly score.

Further we'll explore what factors give higher friendly score for a certain major.

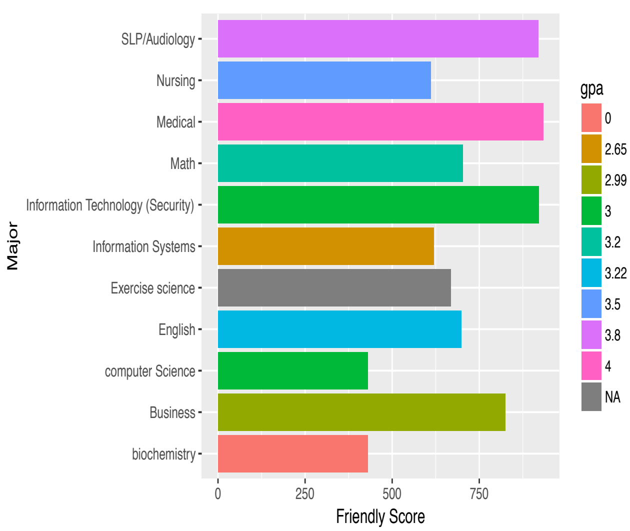

Figure 2: GPA vs Friendly score

3.3 GPA

As shown in figure 1 under same discipline such as IT different majors have varying friendly score so maybe GPA plays a role in higher friendly score. Usually student's with higher GPA have higher prospects of getting jobs, funding etc. Based on above fact one would assume that higher GPA would result in higher score but figure 2 shows different results. Majors such as Information Technology (Security) and Computer science have same GPA 3.0 but their friendly score is very different. That implies GPA has no role in friendly score calculation.

3.4 Number of Friends on Social Media

A large number of friends represents a good social capital. Since friendly score is calculated based on social network data, it implies having a lot of friends would probably give higher social score.

Figure 3: Friends vs Friendly score

Figure 3 shows that student with Math major has highest friends score (40) but still her social score is very less comparatively. Thus even friends score does not seem to be a major factor in the student's social score.

3.5 Number of sources of data

As we know friendly score API offers students to connect their social media profiles from six different sources, could having strong social presence i.e. be on all platforms can be indicative of higher score.

Figure 4 does show some sort of relationship between number of sources of data student has provided and friendly scores.

Information System and Medical majors has higher friendly score and data from four different sources. But at the same time biochemistry has data from 5 sources but friendly

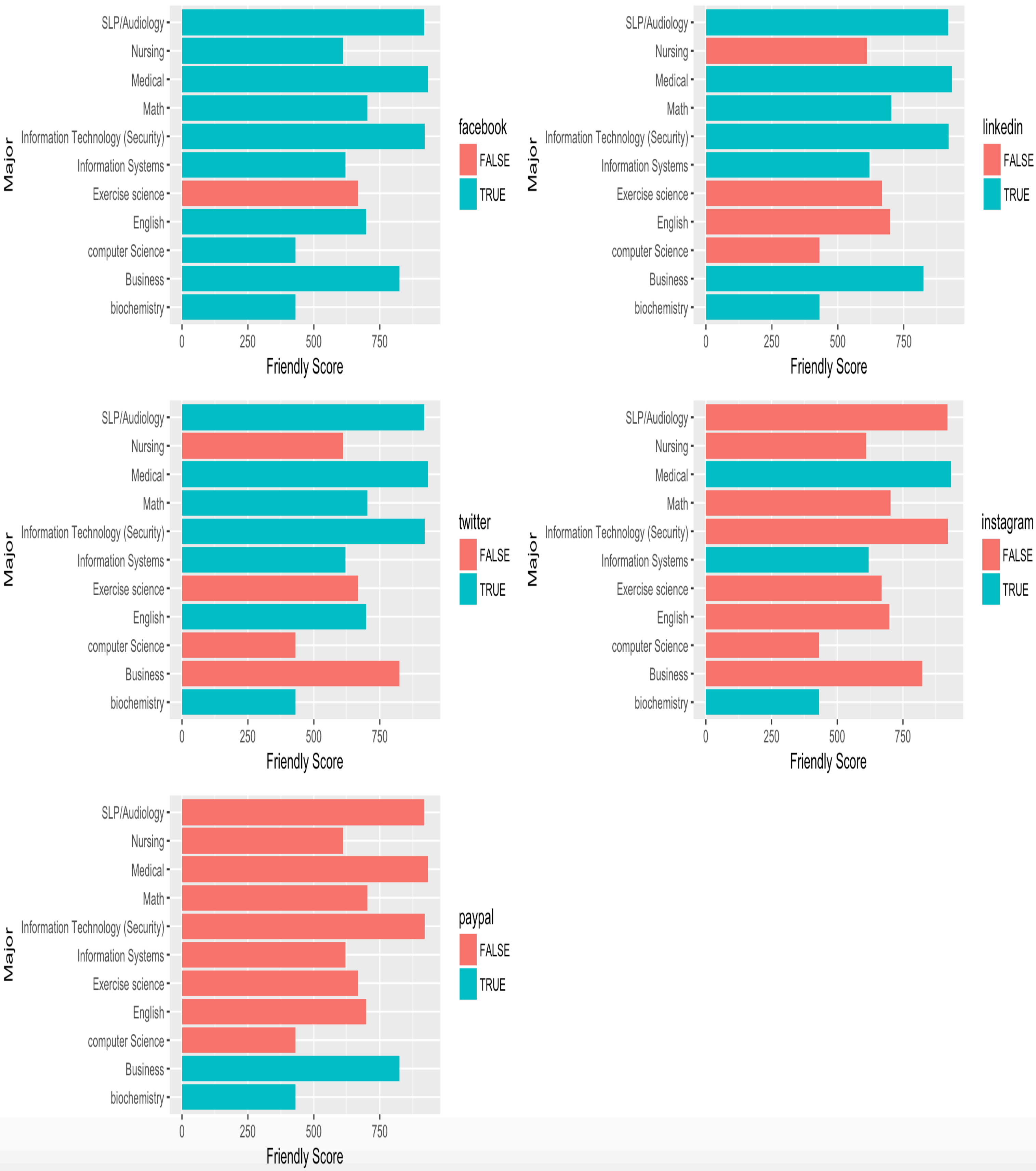

Figure 5: All sources vs Friendly score

score is comparatively less. Based on this it can be assumed that number of sources from which data is collected does matter but is not the sole contributor for social score. Intuitively it makes sense, the more social capital a person has, the higher its social score will be.

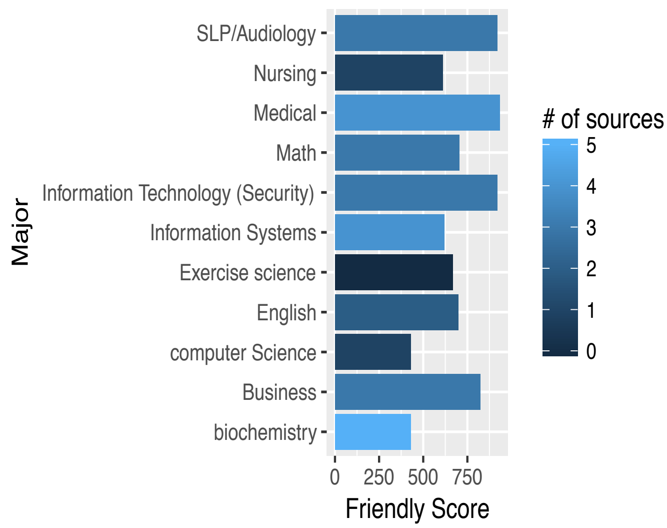

Figure 4: No of sources vs Friendly score

As we see in figure 4 there is some correlation between number of sources of data and friendly score, is it possible that one platform is contributing to higher score compared to others?
Figure 5 shows that it's not dependent on one particular source. It's the cumulative social presence of the student that matters in calculating friendly score. The more profiles student shares with friendly score, chances of good social score are higher.

3.6 Employer

Friendly score API connects to LinkedIn where people post information about their employees, skills, education etc. Could being employed boost one's social score because if you are employed chances of returning loan are higher. Banks or lenders would be more interested in giving loans to such individuals (this is comparable to FICO system).

Figure 6 shows that current dataset does not have the employment information of all students. Most of them have missing value for employer. The ones who have employer information available have higher social score compared to others. The ones who are not employed have relatively lower social score.

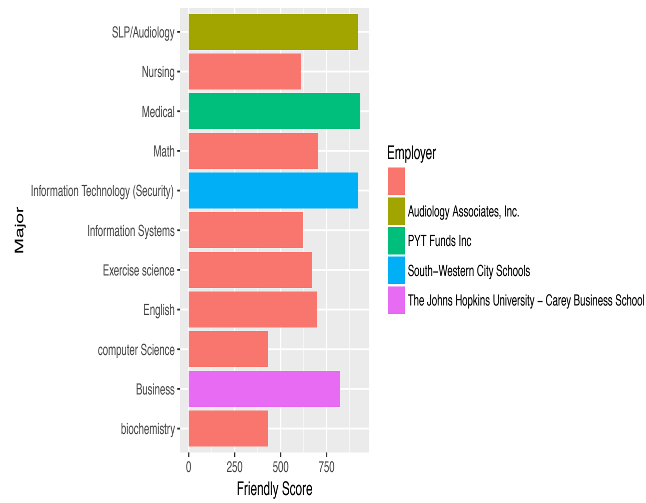

Figure 6: Employer vs Friendly score

3.7 Tabular View of Major's with Higher Score

Let's analyze cumulatively which factors that are analyzed in this paper (major, friend's score, GPA, number of sources of data, employer) are main contributing factors based on above analysis.

Table 1: Cumulative view of factors and Friendly score

Major	Friends Score	GPA	# of Sources of data	Employer	Is Social Score Higher?
SLP/Audiology	No	3.8	4	Yes	Yes
Medical	Yes	4	5	Yes	Yes
Information Security	No	3	4	Yes	Yes
Business	No	2.99	4	Yes	Yes
Math	Yes	3.2	3	No	No
Computer Science	No	3	1	No	No
Nursing	No	3.5	1	No	No
Exercise science	No	NA	0	No	No

As Table 1 shows that students with majors that have higher friendly scores has data from many sources and also employed. At the same time due to missing and less data available it's not strongly conclusive but it does make sense from lender perspective. Strong social presence and being employed are signs that person has social capital and their information can be verified and makes them reasonable candidates for loan approval.

4 Ethical Consequences of Data Collected by Friendly Score

Friendly score's API stores student's data including PII (Personally Identifiable Information) on its secure servers. When lenders request student's score, data can be returned in two formats:

- aggregated score for each social platform (scores only)
- details of students on each platform (including name, email, address etc.)

In former case lenders just gets the social score of student and don't receive any PII data. But in latter case, the payload contains PII data among other information.

When students give permission to Friendly score API to access their social information, it's privacy policy does not explicitly states that its giving loads of information about student to its lenders; it's not just the social score. The question is even if student has signed privacy agreement with friendly score API, does it imply that Friendly score API can return PII data to lenders? What if data breach happens at lender's side [8]?

Another thing accessed by Friendly score is student's friend's information, isn't it violation of friend's privacy because no one asked them in writing if they are okay if their information is accessed.

Another thing worth noting is that there is no expiration date of data retention by Friendly score API and lender. Even if Friendly score's servers are secure; Are they any rules that Friendly score is abiding by to protect user's privacy.

The recommendation is that friendly score API

- comply with EU General Data Protection Regulation (GDPR) Agency
- limit PII data sharing to lenders
- implement data retention policy. It should not be pulling user's data forever. If 'x' years or months has passed, data should be automatically deleted or archived from its servers so in case of data breach it's not available.

5 CONCLUSIONS

In summary from borrowers perspective social score is pretty good alternative to FICO score. It can provide lending opportunities to people who don't have traditional credit history but have strong social presence. But at the same time, it raises ethical concerns around user's data privacy.

The analyses done on different data points shows that the stronger social presence a person has on social media websites it would have higher friendly score. Though it is inconclusive if presence of protected categories affects the social score or not because current dataset has missing values for these attributes. GPA of student does affect to some extent but this data is not collected by Friendly score API so it does not have any role in social score. The number of friends on social media and sources of social media also does not affect social score. It's actually the collective information from all sources that contribute to higher social score. Also its noted that if a student is employed (that means it possess skills to acquire job), it has higher friendly score. Based on above analysis it is concluded that friendly score API does not seem discriminatory. If a student has strong/large social capital chances are it's going to present higher social score thus increasing the lending opportunities for students.

But student's social circle is not always higher. The students living in less fortunate areas might not be employed, in that case their friendly score is going to be low thus less lending opportunities.

Another thing to consider is that social network companies such as facebook, twitter etc has what sort of agreement with companies such as Friendly score or Lenddo? Are they even aware that data is collected from their websites and is used to make credit loan decision? If these systems go mainstream do they have the ability to explain why a certain loan application was accepted or rejection based on social score.

The future work can be to analyze different social network scoring systems to see how they corelate. Ideally for a person 'x' both Friendly Score and Lenddo should give same social score.

ACKNOWLEDGMENTS

This work is done for PYTFund to assess the feasibility of friendly score API for social scoring of students. Based on this social score PYTFund will lend money to students.

REFERENCES

[1] https://rctom.hbs.org/submission/its-all-about-who-you-know-lenddo-makes-credit-decisions-based-on-your-social-network/
[2] https://www.cnbc.com/2015/05/05/credit-invisible-26-million-have-no-credit-score.html
[3] Lenddo: https://www.lenddo.com/
[4] Friendlyscore : https://friendlyscore.com/
[5] PYTFund: http://www.gopyt.com/
[6] DeDeo, Simon. "Wrong side of the tracks: Big Data and Protected Categories." (2016) DOI: https://arxiv.org/abs/1412.4643
[7] GDPR: https://www.eugdpr.org/
[8] Cambridge Analytica : https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html