Machine Learning System

Duration: 20 Minutes
Difficulty: Medium
Domains: Fraud

Problem

Given the following attributes of Facebook's user data listed below, come up with 25 features that can help identify whether a school name on a Facebook's user profile is fake.

User Profile	Backend Data	Graph Data
1. Name	1. User's date of sign up	1. Friends of a user
2. Email address	2. IP Address	2. Community membership
3. School name		
4. Hometown	Event Data	
5. Current residence	Friend request sent	
6. Birth date	Friend request received	
7. Relationship status	Friend request accepted	
8. Interests	Posts	
9. Gender	Status Updates	

Solution

[Candidate] To answer this question, I want to brainstorm why bad actors would create fake school profiles. Understanding the motives helps extract features.

[Interviewer] Sounds good.

[Candidate] I see two primary reasons that bad actors create accounts with fake schools on Facebook: (1) spamming and (2) predatory behavior. Based on the motives and data available, we can define the features.

[Interviewer] Okay, let's talk about features.

[Candidate] The quickiest way to derive an ample of features is to brainstorm sets of features. The first feature set comes from the school name variable:

Feature 1 - The number of occurrences of the school name.

The higher the occurrences of the school name across users, the more likely that the school is real.

Feature 2 - The number of character repeats (i.e. eeeea = 4).

Feature 3 - Consonant-vowel ratios (i.e. Brocks = 5 / 1).

Both features 2 and 3 gauge whether a name is spammy as in "aaaaa high school."

[Interviewer] Sounds good. What else can you think of?

[Candidate] I could derive a feature set on the profiles.

Feature 4 - Gender Indicator

Though the following hypothesis should be tested. I presume that the bulk of the child predators on Facebook are older men pretending to be teenage boys so they can target teenage girls.

Feature 5 - Relationship Indicator

I would presume that sexual predators would pretend to be single or no status to allure single teenage girls.

Feature 6 - The number of occurrences of the hometown.

Similar to feature 1, I would apply the same analysis on feature 6, extract the number of occurrences.

Feature 7 - Email handle character repeats (i.e. ieeeee@gmail.com = 5)

Bad actors would use spam-like email addresses to create fake accounts. The character analysis scores on email handles assesses spamminess.

[Interviewer] Okay, sounds good. Can you think of any features derived from the IP address field?

[Candidate] Can I presume that there's a blacklist of IP address?

[Interviewer] Yes. But, in this exercise, let's presume that the list is not available.

[Candidate] I see. I can think of an alternative approach to creating features using the IP address field. I believe it is in the interest of bad actors to create as many fake accounts to target multiple users simultaneously. This means that multiple accounts could be tied to a single IP address. I could see a handful of features derived from this idea.

Feature 8 - The number of distinct email address per IP address the past 1 day

Feature 9 - The number of distinct email address per IP address the past 7 days

Feature 10 - The number of distinct email address per IP address the past 30 days

Feature 11 - The number of distinct email address per IP address lifetime

An organic users would often have one set of PII's and IP address per account. Bad actors could have multiple. This approach can be applied on other PII's.

Feature 12 - The number of distinct full names per email address the past 1 day

Feature 13 - The number of distinct full names per email address the past 7 days

Feature 14 - The number of distinct full names per email address the past 30 days

Feature 15 - The number of distinct full names per email address lifetime

Feature 16 - The number of distinct birthdates per email address the past 1 day

Feature 17 - The number of distinct birthdates per email address the past 7 days

Feature 18 - The number of distinct birthdates per email address the past 30 days

Feature 19 - The number of distinct birthdates per email address lifetime

[Interviewer] Okay sounds good. How can you use other variables from the event and connection data sources?

[Candidate] Similar to feature engineering using the IP address field, deriving features should help distinguish organic to non-organic behaviors. I can apply this idea to create these features:

Feature 20 - The friend requests to friend received ratio. I would assume that sexual predators are more likely to send friend requests than receive them. Organic users are less inclined to send friend requests to strangers.

Feature 21 - Number of friend requests sent the past 30 days

Similar to the window-based features using the IP address field, the distribution of on the volume of friend requests sent is most likely higher for the bad actors than organic users

Feature 22 - The number of posts the past 30 days

I would anticipate that fake profile users would generate less posts than organic users.

Feature 23 - The number of distinct IP addresses to friends ratio.

It might be possible that the bad actor may attempt to create an authentic profile by adding friends whose profiles are also fake. The ratio could be lower for bad actors than organic users.

Feature 24 - # of friends

Fake users would generally have less friends than organic users.

Feature 25 - Betweenness score

In statistical network analysis, there are measures to assess how well an individual is connected among other users. The most connected the user, the higher the betweeness score. Organic users would have higher betweeness score than fake users.

Interviewer Solution

To ace this question:

1. This is a question designed to assess your ability to perform feature engineering. You will be assessed based on whether you could conjure up 30 features for a model that address the business problem of flagging fake school names. Additionally, you will be assessed based on the quality of your reasoning behind the features.
2. To approach this problem, you need to evaluate the business problem. *Why does Facebook care whether the school name is legitimate or not? What kind of users would use fake school names?* When you answered these questions, putting yourself in the shoes of such users will help you come up with features that can help the model differentiate between good users and bad users.
3. An effective strategy to come up with a lot of features is to think about the features in terms of categories. For instance, using graph data, you can come up with several features that are related to a user's network.

Before you proceed with listing features to solve the problem of identifying fake schools on Facebook profiles. You need to first delve into the problem.

Understanding the problem

Facebook wants to ensure that the quality experience of their platform is not compromised due to fake users – the kind of users that would want to mask their identity with fake profile details including school names. The motivation behind such users could be the following:

1. Cyberbullying.
2. Blackmailing.
3. Sexual predators targeting an underage person.
4. Trolling or pranking normal users.
5. Spamming advertisements via spam bots.
6. Disseminating fake news.

Feature Engineering

To approach this feature engineering exercise, you need to put yourself in the mind of the bad users who create profiles with fake schools. *What kind of behaviors would such users possess that differentiate them from normal users?*

Feature Categories	Features	Reasonings
Character Analysis	1. Consonant-to-vowel ratio	Applied on text-based variables, such as school names and email addresses, the ratios can help highlight spam-like names of email addresses such as *aaaa.bbdd@gmail.com* and school names such as *Rsaidkes Elementary School*.
	2. Character-repeat ratio	
	3. Keyboard distance ratio	
Profiles	4. Age	The features measure realness of a user profile. One assumption of bad users is that their profiles would provide personal information than those of ordinary users. To project an authentic image and raise social proof, ordinary users would craft a profile enriched with personal details and photos. Malicious users, on the other hand, face limits
	5. Gender	
	6. Profile photo missing indicator	
	7. Number of photos	
	8. Number of interests	
	9. Relationship status	
	10. Hometown missing indicator	
	11. Current residence missing indicator	

		when providing details and photos that achieve authenticity.
Network	12. # of degrees 13. Betweenness centrality 14. Closeness centrality 15. Community network density	The features are metrics showing how well-connected a user is to a network of other users. Facebook is merely a reflection on cliques and communities users are part of in real-life. A network of friends or cliques in the same school would be well-connected among themselves. However, malicious users with fake school names are less likely to be in a well-connected of friends on Facebook.
Events	16. # of status updates 17. # of posts 18. # of friend requests sent 19. # of friend requests accepted 20. # of friend requests received	The features measure how active th
IP	21. # of users per IP during the past 1 week 22. # of users per IP during the past 4 week 23. # of users per IP during the past 8 week 24. # of users per IP during the past 16 week 25. # of users per IP during the past 32 week	The bad actor may possess multiple users to conduct their goals in harming as many ordinary users as possible. Some of the users could be flagged with existing fraud model or reported by current users. Creating multiple users is a way to address the possibility of a ban. Aggregating the number of users per IP is a way to identify actors with this behavior of creating several profiles.
Email	26. # of users per email during the past 1 week 27. # of users per email during the past 4 week 28. # of users per email during the past 8 week 29. # of users per email during the past 16 week 30. # of users per email during the past 32 week	Similar to aggregating across IPs, emails aim to achieve similar goal in identifying actors with malicious intents conducted via several fake profiles.

Interviewer Assessment

In the statistics section, a candidate is assessed based on correctness and soundness of statistical methodology, product sense and communication. For each dimension the candidate is rated in the following scale: (5) superior, (4) good, (3) adequate, (2) marginal, (1) not competent.

Assessments	Rating	Comments
ML Methodology	5	The candidate completed the problem with 25 features that could work in identifying users with fake schools. Rather than creating one feature per variable, she shared ways to create a handful when combined with other variables. For instance, she applied aggregations on the IP address field to count distinct PII field on a moving window. This effective strategy alone helped her create ⅓ of her feature set.
Product Sense	5	She put herself in the shoes of a bad actor to extract features that could classify whether a school is fake. She derived features based on the premise that bad actors would use fake schools for spamming and predatory behaviors. With this premise, she succeeded in deriving 25 features.
Communication	5	Her approach to her solution was well-structured. She did not dive immediately into a solution. She warmed-up with a few assumptions about bad actors. This helped her brainstorm features. As she listed her features, she explained the reasonings for them.