

Identifying Fake Job Offers Online

Big Data for Social Sciences | Shir Weinbrand

Motivation and Research Question



01 During Covid-19 - one of the biggest fraud waves

Submitting fake claims for unemployment benefits, through 'identity theft'.

This is how approximately 7 billion dollars were stolen, 800 million from the state of New York alone.

<https://www.propublica.org>

02 How is identity stolen?

Identity thieves post fake job ads online   

Job seekers fill in their personal details (+ CV)

The identity thieves steal the details and submit a claim for unemployment benefits

<https://mybackgroundcheck.com/news/identity-theft-using-fake-jos>

03 The research question

Is it possible to identify fraudulent job offers online, and thereby prevent job seekers from falling victim to identity theft scams?

Examples of Fake Ads

The image is a screenshot of a WhatsApp chat interface. At the top, there is a header for 'Yahoo Boys Community' with a 'SCAM' warning icon and '5,160 members'. Below this, a 'Pinned message #1' is shown with the text '-----FORMAT----- - Hello, *Greetings* - Your primary duties include p'. The main chat area has a green background with a pattern of small icons. A message from 'Documents' is visible, dated 'July 25', with a right arrow indicating it's a video. Below this, a white card contains the text 'Cash app for sell \$50' and 'I accept btc only'. A redacted area follows. Then, a blue link for 'WhatsApp.com' is shown, followed by the text 'Open WhatsApp' and 'WhatsApp Messenger: More than 2 billion people in over 180 countries use WhatsApp to stay in touch with friends and family, anytime and any...'. To the right of this text is the WhatsApp logo. At the bottom of the card, it says '1064' and '4:44 PM'. Below the card, another message is partially visible: '-----FORMAT----- - Hello, *Greetings* - Your primary duties include picking up passengers, providing on-time shuttle service assisting passengers by answering question, providing directions or attending to their other needs. This includes helping passengers with inquiries about delayed or lost baggage'. Further down, it says 'Mon - Friday', '7 hours per day', '\$2000 per week', and 'I will need more information from you to proceed to our interview section.. Reply with a message to proceed.'.

-

I would be asking some questions now, To know if you qualify for the application before proceeding to the verification stage.

- Do you have a customer service experience?
- Do you have excellent communication skills ?
- Do you Have a Friendly and Affable personality ?
- Do you have a clean Driving record ?

—

- Congratulations, you are qualified based on the answers given and your resume , now we would be proceeding to the verification process..

- Your Identity would be verified by the Government trusted digital verification system .. "ID.me" to make sure you're you and not someone pretending to be you , upon completion of identity verification you would receive an email from Shuttle DDBVA within 1-2 business day..


- We would need the following DETAILS to proceed with the identification process

FIRST NAME
LAST NAME
PHONE NUMBER
EMAIL ADDRESS
SSN
PICTURE OF DRIVERS LICENSE FRONT AND BACK

THAT'S ALL, Once the client gives you the details, buzz me on WhatsApp and let start work on it Asap.


facebook

Log In



Trucking Jobs, Trucking Companies and Trucking News

Join Group



September 17 at 7:46 PM · 🌐

Hello ,
primary duties include picking up passengers, providing on-time shuttle service assisting passengers by answering question, providing directions or attending to their other needs. This includes helping passengers with inquiries about delayed or lost baggage
Mon - Friday
7 hours per day
\$2000 per week


6 Comments

Like

Comment


Share

View 1 more comment




Location?

Share · 3w



Location, where can we apply?

Share · 1w



Must be spam

Share · 1w

Methodology

01

data

Using a dataset from the Aegean Employment Fraud Department (EMSCAD, downloaded from Kaggle). Contains 17,880 job ads, of which 17,014 are genuine and 866 are fake.

Key variables:

- description (job description)
- fraudulent (real (0)/fake (1))
- required_education (level of education)
- salary (was salary specified)
- has_logo (appearance of the company logo)

02.....

Classification - Fake VS. Real Supervised learning

Using the *Naive Bayes* classifier to predict a real ad versus a fake one. Comparing the result to logistic regression, while creating dummy variables and using the variables:

- education level
- Salary specified/not
- A logo appears/no

03.....

Text analysis - Topic Modeling Unsupervised learning

Using *LDA* to identify key concepts related to fake ads.

- *Word Cloud* to illustrate the key concepts.

job_id	title	location	department	salary_range	salary	company_profile	description	requirements	benefits	lecommuti	has_company_logo	s_questio	employment_tred_exper	required_education	industry	function	fraudulent
1	arketing Inte, NY, New Y		Marketing		1	most talented hit systemsRes	ulti-tasker and juggler of respc			0	1	0	Other Internship			Marketing	0
2	se - Cloud VZ, , Aucklar		Success		1	video producti	delivery of videmputer knowledge for this rol			0	1	0	Full-time ot Applicable			Marketing and Advstomer Sen	0
3	Machinery S, IA, Wever				1	skills, experiendetail. A strong	prepare forms for all pending t			0	1	0					0
4	cutive - WaDC, Washir		Sales		1	of the changinss world., Priv	and advocacy witherous outdc			0	1	0	Full-time d-Senior lev	Bachelor's Degree	puter Softv	Sales	0
5	review Man FL, Fort Worth				1	orary-to-hire, ations.DUTIES d, Self-StarterExc	3enefits Off			0	1	1	Full-time d-Senior lev	Bachelor's Degree	tal & Health:h Care Pro		0

Results

Naive Bayes

- Accuracy - 95.25%
- Sensitivity (Fake when fake) - 42.515%
- Specificity (Real when real) - 97.82%
- FP (predict fake, in real - real) - 2.17%
- FN (predict real, in real - fake) - 57.48%

Prediction	Reference/True	
	(Not Fake) 0	(Fake) 1
(Not Fake) 0	3335	96
(Fake) 1	74	71

Error Rate = 0.0475 (4.75%)

FP = 0.0217 (2.17%)

FN = 0.57485 (57.48%)

Logistic Regression

- Accuracy - 89.26%
- Sensitivity (Fake when fake) - 21.858%
- Specificity (Real when real) - 92.897%
- FP (predict fake, in real - real) - 7.10%
- FN (predict real, in real - fake) - 78.14%

Prediction	Reference/True	
	(Not Fake) 0	(Fake) 1
(Not Fake) 0	3152	143
(Fake) 1	241	40

Error Rate = 0.1074 (10.74%)

FP = 0.07103 (7.10%)

FN = 0.78142 (78.14%)

	Accuracy	Error Rate	False Positive	False Negative
NB	95.25%	4.75%	2.17%	57.48%
LR	89.26%	10.74%	7.10%	78.14%

Results

LDA

LDA - Fraudulent Jobs				
	Requirements	Temptation/Extra Benefit	Terms of Employment	Areas of Responsibility/Industries
1	customer	food	time	project
2	experience	free	per	solutions
3	product	looking	looking	equipment
4	team	skills	home	engineering
5	data	luxury	position	within
6	management	board	hours	position
7	business	ultra	experience	support
8	company	cruise	money	maintenance
9	service	service	online	systems
10	skills	help	week	oil

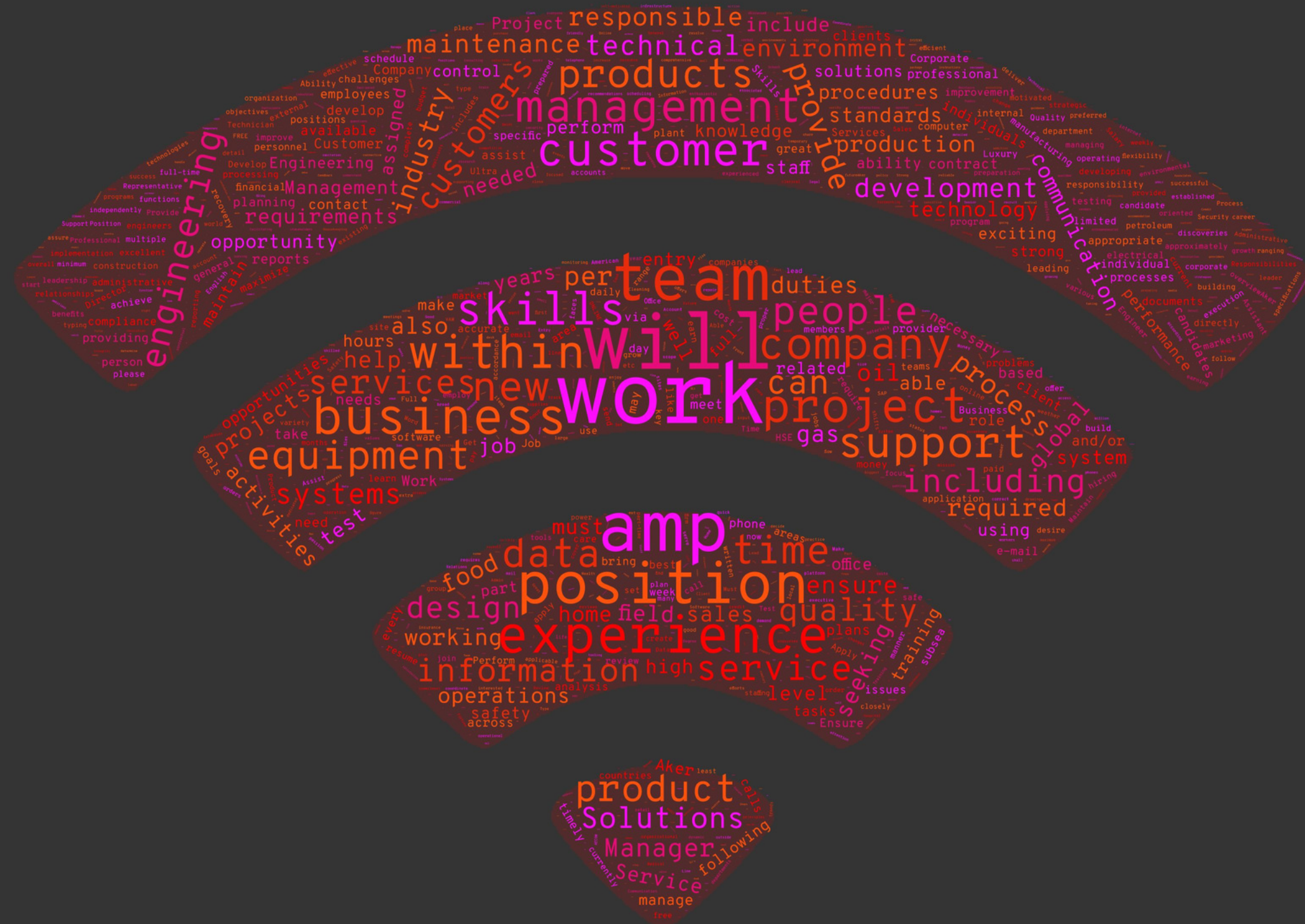
doc_id	text	title	salary_range	required_experience	required_education	industry	fraudulent	topic
358	If you have experience in financing for auto sales and a great attitude you can work in our Hazelcrest office. From \$500 top \$1000 a week by contract.	Financing Auto(car) sales	35000-73000	Associate		Automotive	1	Temptation/Extra Benefit
145	The group has raised a fund for the purchase of homes in the Southeast. The student on this project will help them build their investments from the ground up and will help with the analysis and modeling of their investments. We should be looking for someone with a strong general finance skills and has a lot of entrepreneurial ability.	Forward Cap.					1	Requirements

Topics Distribution - Theta				
	Requirements	Temptation/Extra Benefit	Terms of Employment	Areas of Responsibility/Industries
358	4%	58%	27%	12%
145	46%	9%	43%	2%

Topics Popularity			
Requirements	Temptation/Extra Benefit	Terms of Employment	Areas of Responsibility/Industries
44%	9%	27%	20%

- Top 10 words in each topic
- Examples of job postings, related topics & Theta
- Topics popularity

Results



Word Cloud

- WIFI-sign
- Popular words in fake jobs, like:
*business, position, experience,
product, solutions, support, project*

Thanks, and...



If it sounds
too good
to be true ...

... it probably is

<https://www.accc.gov.au/publications/if-it-sounds-too-good-to-be-true>