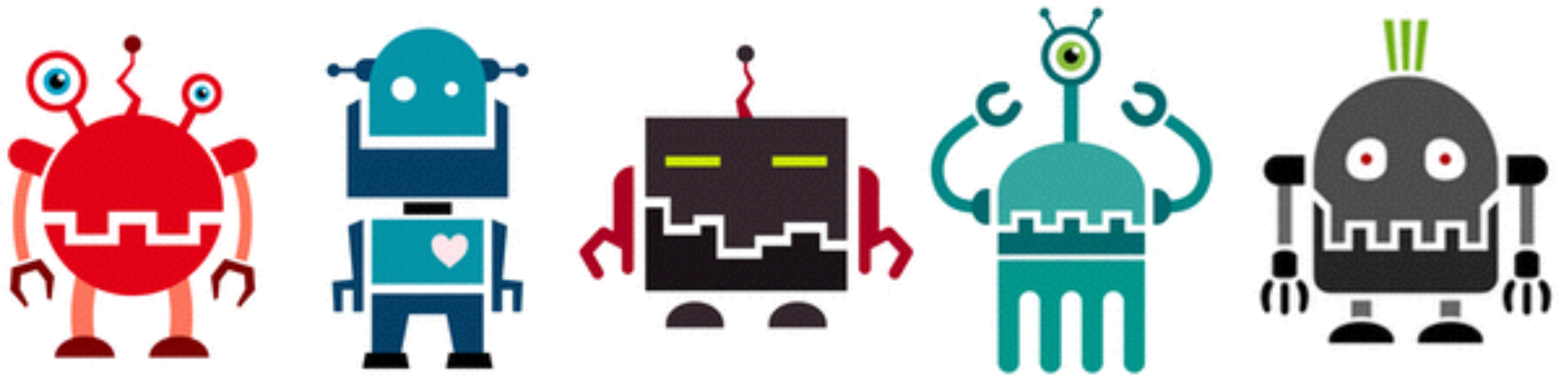


Facebook Recruiting IV:

Human or Robot?

Tucker Silva



The goal is to identify online auction bids that are placed by "robots", helping the site owners easily flag these users for removal from their site to prevent unfair auction activity.

Predict if an online bid is made by a machine or a human

Bidder Data

bidder_id - Unique identifier of a bidder.

payment_account - Payment account associated with a bidder. These are obfuscated to protect privacy.

address - Mailing address of a bidder. These are obfuscated to protect privacy.

outcome - Label of a bidder indicating whether or not it is a robot. Value 1.0 indicates a robot, where value 0.0 indicates human.

Bids Data

bid_id - unique id for this bid

bidder_id - Unique identifier of a bidder (same as the bidder_id used in train.csv and test.csv)

auction - Unique identifier of an auction

merchandise - The category of the auction site campaign, which means the bidder might come to this site by way of searching for "home goods" but ended up bidding for "sporting goods" - and that leads to this field being "home goods". This categorical field could be a search term, or online advertisement.

device - Phone model of a visitor

time - Time that the bid is made (transformed to protect privacy).

country - The country that the IP belongs to

ip - IP address of a bidder (obfuscated to protect privacy).

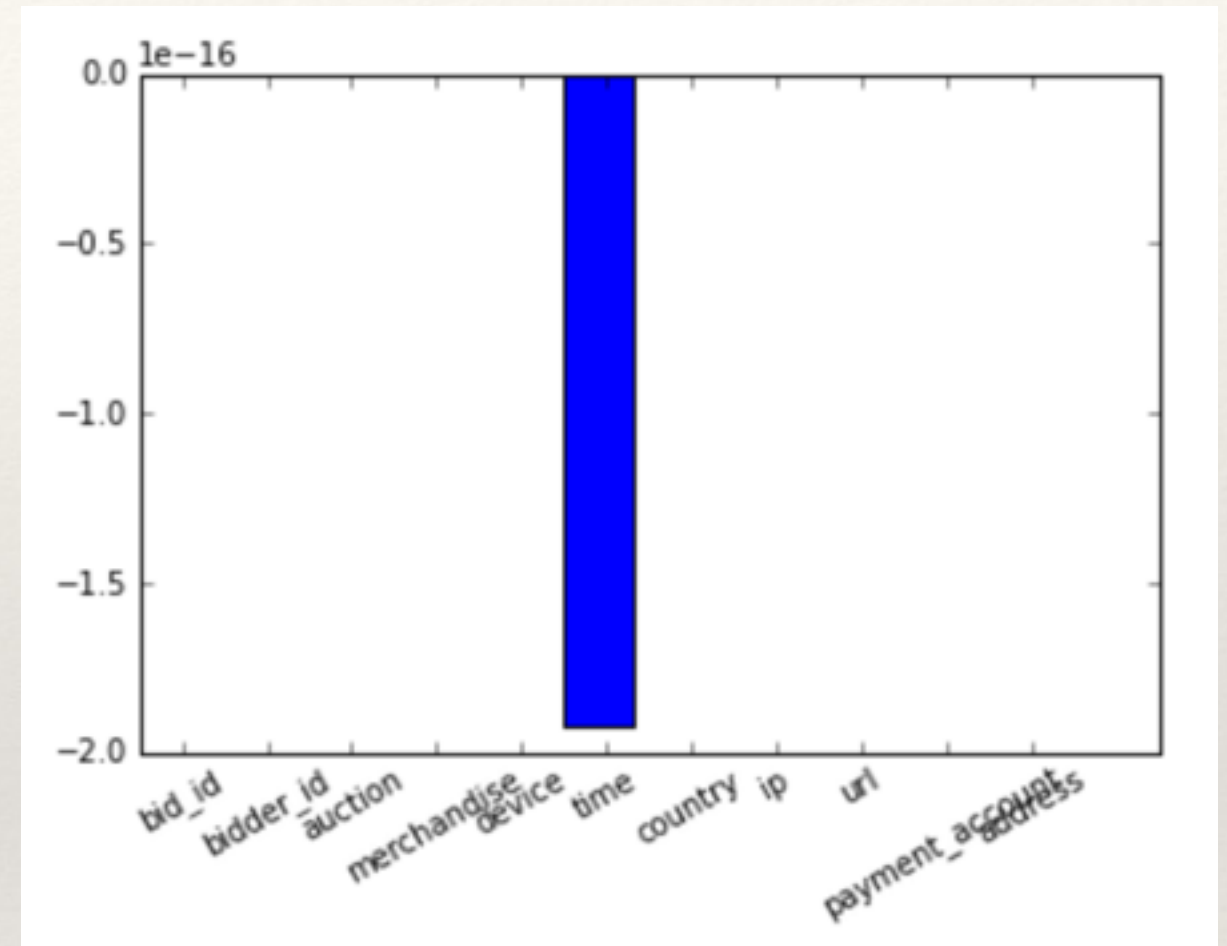
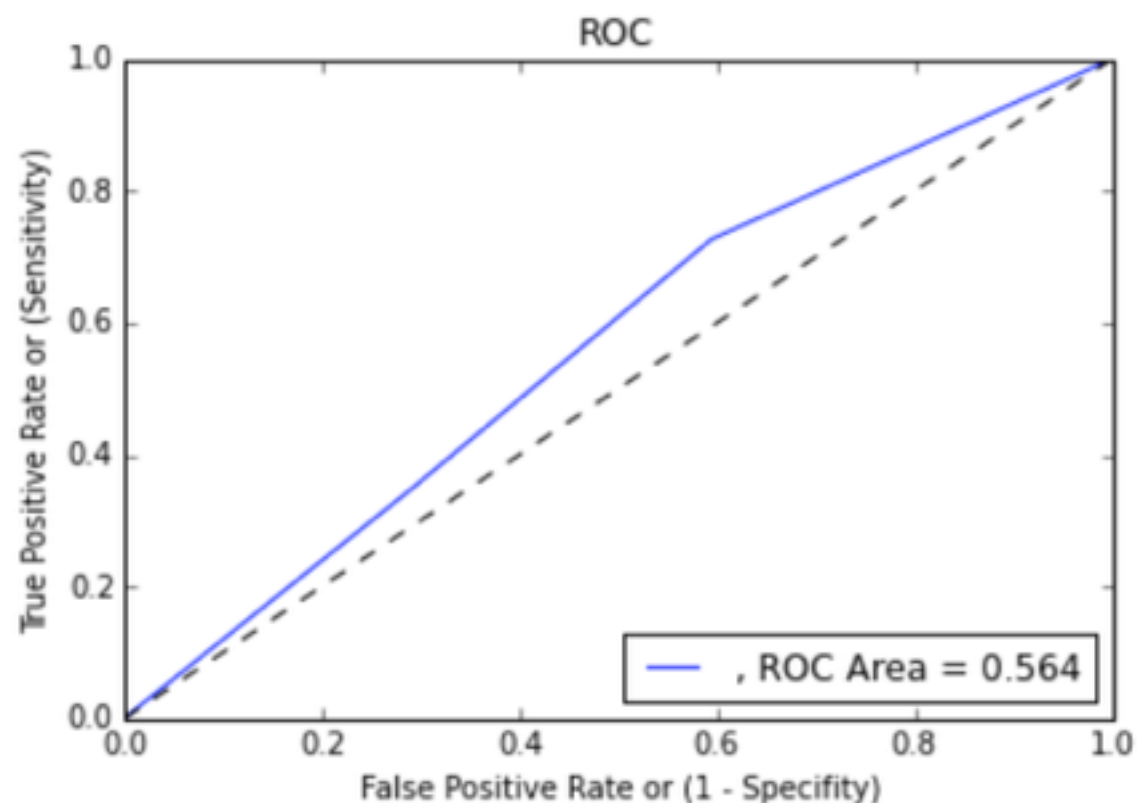
url - url where the bidder was referred from (obfuscated to protect privacy).

	bidder_id	payment_account	address	outcome
0	91a3c57b13234af24875c56fb7e2b2f4rb56a	a3d2de7675556553a5f08e4c88d2c228754av	a3d2de7675556553a5f08e4c88d2c228vt0u4	0
1	624f258b49e77713fc34034560f93fb3hu3jo	a3d2de7675556553a5f08e4c88d2c228v1sga	ae87054e5a97a8f840a3991d12611fdcrfbq3	0
2	1c5f4fc669099bfbfac515cd26997bd12ruaj	a3d2de7675556553a5f08e4c88d2c2280cybl	92520288b50f03907041887884ba49c0cl0pd	0
3	4bee9aba2abda51bf43d639013d6efe12iycd	51d80e233f7b6a7dfdee484a3c120f3b2ita8	4cb9717c8ad7e88a9a284989dd79b98dbevyi	0
4	4ab12bc61c82ddd9c2d65e60555808acqgos1	a3d2de7675556553a5f08e4c88d2c22857ddh	2a96c3ce94b3be921e0296097b88b56a7x1ji	0

	bid_id	bidder_id	auction	merchandise	device	time	country	ip	url
0	0	8dac2b259fd1c6d1120e519fb1ac14fbqvax8	ewmzr	jewelry	phone0	9759243157894736	us	69.166.231.58	vasstdc27m7nks6
1	1	668d393e858e8126275433046bbd35c6tywop	aeqok	furniture	phone1	9759243157894736	in	50.201.125.84	jmqhlhflrzwuay9c
2	2	aa5f360084278b35d746fa6af3a7a1a5ra3xe	wa00e	home goods	phone2	9759243157894736	py	112.54.208.157	vasstdc27m7nks6
3	3	3939ac3ef7d472a59a9c5f893dd3e39fh9ofi	jefix	jewelry	phone4	9759243157894736	in	18.99.175.133	vasstdc27m7nks6
4	4	8393c48eaf4b8fa96886edc7cf27b372dsibi	jefix	jewelry	phone5	9759243157894736	in	145.138.5.37	vasstdc27m7nks6
5	5	e8291466de91b0eb4e1515143c7f74dexy2yr	3vi4t	mobile	phone7	9759243157894736	ru	91.107.221.27	vasstdc27m7nks6
6	6	eef4c687daf977f64fc1d08675c44444raj3s	kjlzx	mobile	phone2	9759243210526315	th	152.235.155.159	j9nl1xmo6fqhcc6
7	7	ab056855c9ca9d36390feae1fa485883issyg	f5f6k	home goods	phone8	9759243210526315	id	3.210.112.183	hnt6hu93a3z1cpc
8	8	d600dc03b11e7d782e1e4dae091b084a1h5ch	h7jjx	home goods	phone9	9759243210526315	th	103.64.157.225	vasstdc27m7nks6
9	9	a58ace8b671a7531c88814bc86b2a34cf0crb	3zpkj	sporting goods	phone4	9759243210526315	za	123.28.123.226	vasstdc27m7nks6

Linear Regression No Class Weight

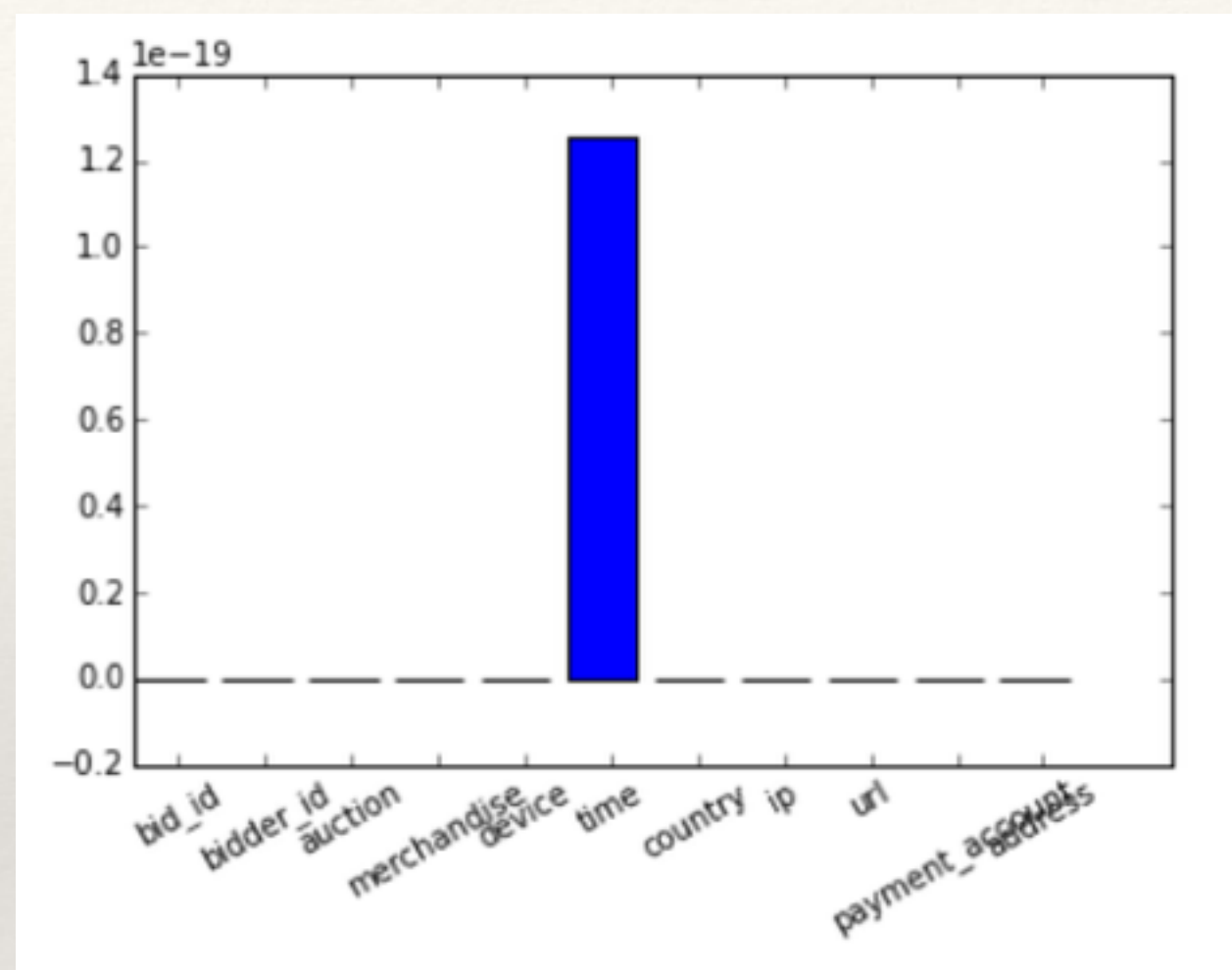
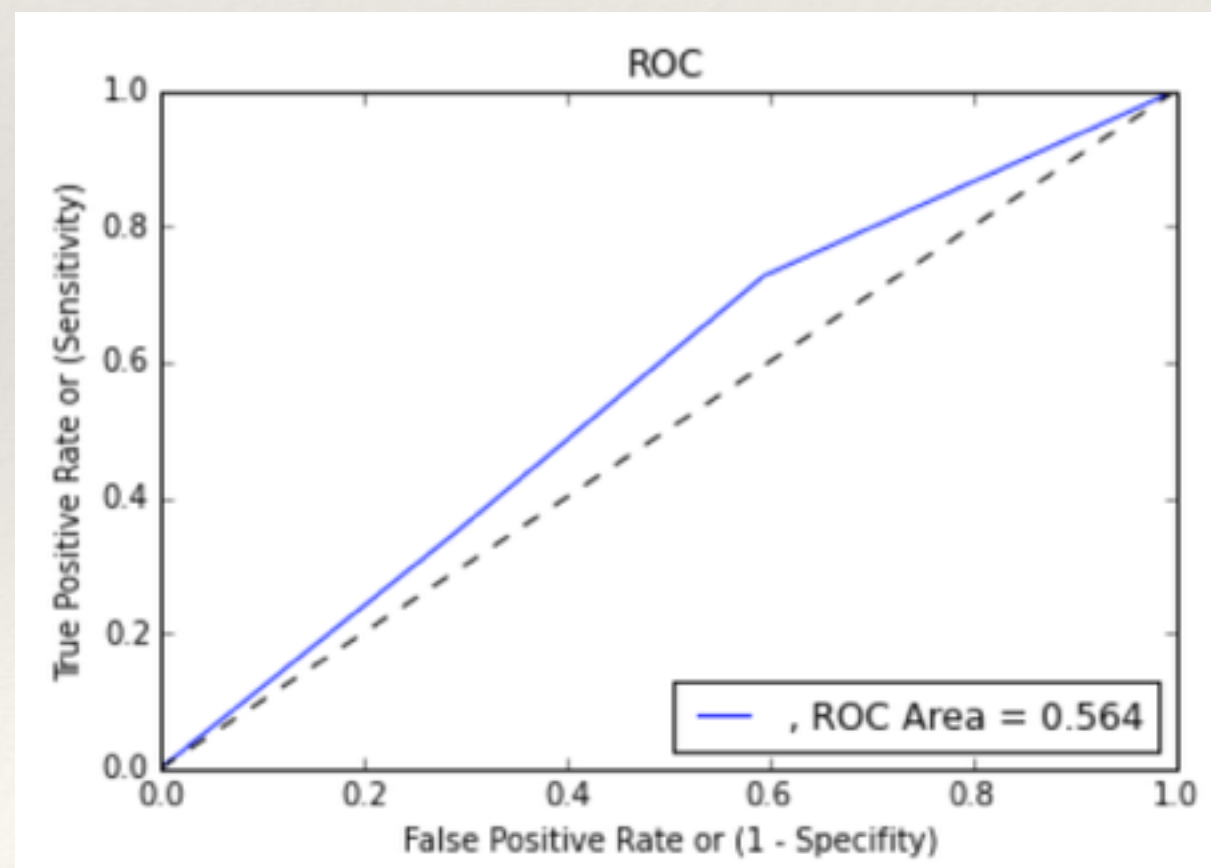
This model is not doing very well at predicting the spam in the train dataset. The overall accuracy is 86.6%, but just guessing "not spam" every time would yield better results and it is not catching any spam.



	precision	recall	f1-score	support
not spam	0.87	1.00	0.93	531670
spam	0.00	0.00	0.00	82575
avg / total	0.75	0.87	0.80	614245

Linear Regression Class_Weight = 'Auto'

This model improved the precision of predicting spam to 13%, however the overall accuracy is only 13% and the precision of predicting "not spam" dramatically decreased to 0.



	precision	recall	f1-score	support
not spam	0.00	0.00	0.00	531670
spam	0.13	1.00	0.24	82575
avg / total	0.02	0.13	0.03	614245

Next