

Identifying the actual PII category

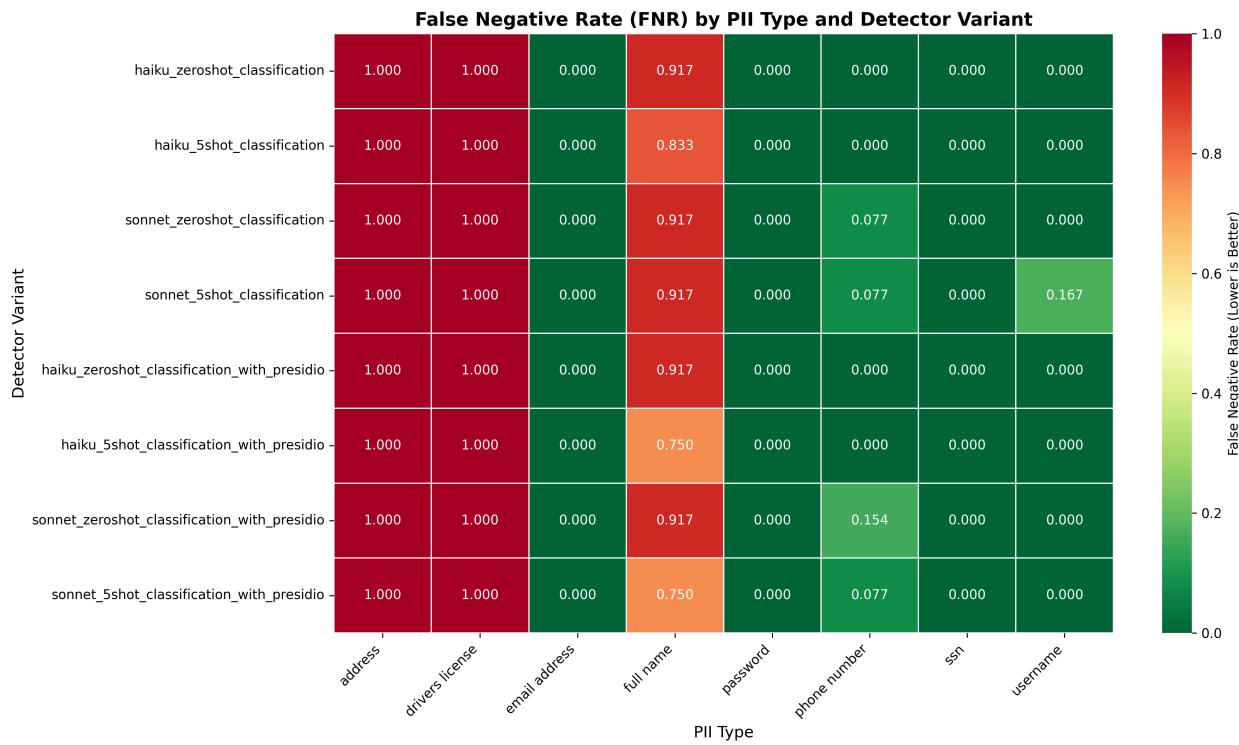
Previously, we tested the ability of LLMs to binary classify emails as containing PII or not. Now we move on to classifying the categories themselves.

Summary

LLMs perform much better in some PII categories than others. Production pipeline should warn client that LLMs might not accurately flag out emails that contain addresses, driver's license and full name PII-combos.

Experiment

We run an experiment to see how each of our variants perform by FNR for each PII type. We found that there are three PII types that have unusable FNR rates: address, drivers license and full name.



Examples

Address, ID 47

Brenda,

I checked my records and I mailed check #1178 for the normal amount on August 28th. I mailed it to 4303 Pate Rd. #29, College Station, TX 77845. I will go ahead and mail you another check. If the first one shows up you can treat the 2nd as payment for October.

I know your concerns about the site plan. I will not proceed without getting the details and getting your approval.

I will find that amortization schedule and send it soon.

Phillip

License, ID 250

Imelda,

The numbers you requested:
139-12-6091
C1154589

thank you,

Phillip

Full name, ID 10

----- Forwarded by Phillip K Allen/HOU/ECT on 10/16/2000
01:42 PM -----

"Buckner, Buck" <buck.buckner@honeywell.com> on 10/12/2000 01:12:21 PM
To: "'Pallen@Enron.com'" <Pallen@enron.com>
CC:
Subject: FW: fixed forward or other Collar floor gas price terms

Phillip,

> As discussed during our phone conversation, In a Parallon 75 microturbine
> power generation deal for a national accounts customer, I am developing a
> proposal to sell power to customer at fixed or collar/floor price. To do
> so I need a corresponding term gas price for same. Microturbine is an

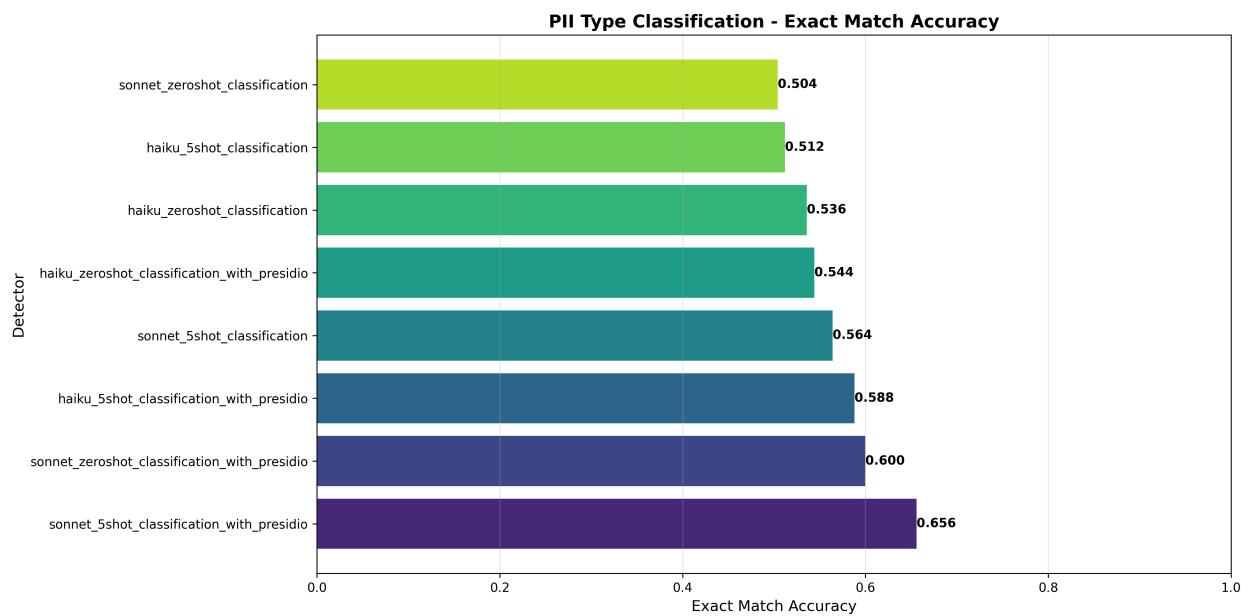
- > onsite generation product developed by Honeywell to generate electricity
- > on customer site (degen). using natural gas. In doing so, I need your
- > best fixed price forward gas price deal for 1, 3, 5, 7 and 10 years for
- > annual/seasonal supply to microturbines to generate fixed kWh for
- > customer. We have the opportunity to sell customer kWh 's using
- > microturbine or sell them turbines themselves. kWh deal must have limited/
- > no risk forward gas price to make deal work. Therein comes Sempra energy
- > gas trading, truly you.
- >
- > We are proposing installing 180 - 240 units across a large number of
- > stores (60-100) in San Diego.
- > Store number varies because of installation hurdles face at small percent.
- >
- > For 6-8 hours a day Microturbine run time:
- > Gas requirement for 180 microturbines 227 - 302 MMcf per year
- > Gas requirement for 240 microturbines 302 - 403 MMcf per year
- >
- > Gas will likely be consumed from May through September, during peak
- > electric period.
- > Gas price required: Burnertip price behind (LDC) San Diego Gas & Electric
- > Need detail breakout of commodity and transport cost (firm or
- > interruptible).
- >
- > Should you have additional questions, give me a call.
- > Let me assure you, this is real deal!!
- >
- > Buck Buckner, P.E., MBA
- > Manager, Business Development and Planning
- > Big Box Retail Sales
- > Honeywell Power Systems, Inc.
- > 8725 Pan American Frwy
- > Albuquerque, NM 87113
- > 505-798-6424
- > 505-798-6050x
- > 505-220-4129

> 888/501-3145

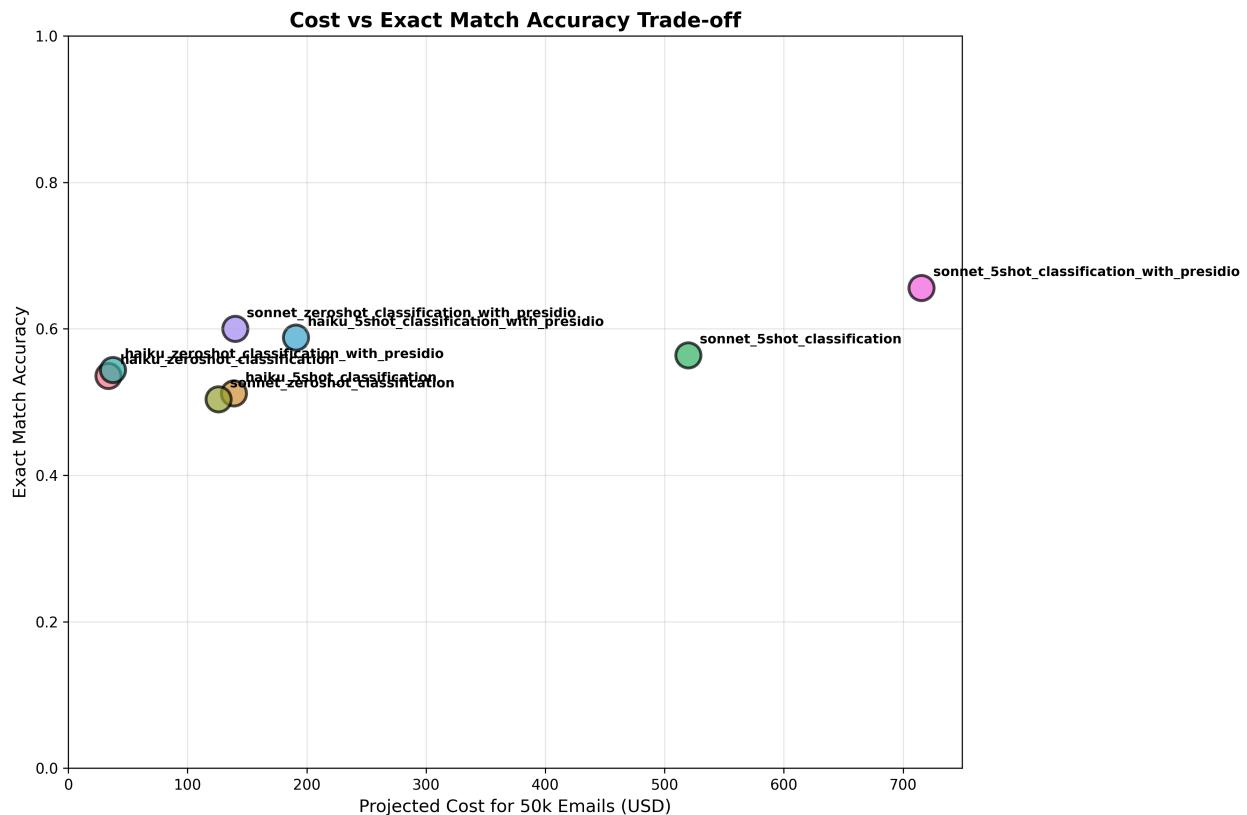
>

Appendix

Exact match accuracy is only at 66%



Cost is inelastic to exact match accuracy



Redder is better. Note that email address has a bad F1 score even though FNR is low as shown above.

