.conf2015

Using Prelert and Splunk Cloud To Fight a Billion Dollar Fraud Problem

Andrew Linn
SVP, Chief Information Security Officer – Orrstown Bank

Christopher Thompson
SVP, Chief Architect – Orrstown Bank

Dr. Steve Dodson
CTO – Prelert

splunk>

Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make.

In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

About ORRSTOWN BANK

Orrstown Bank is a **community bank in PA and MD** with about $1.2 billion in assets

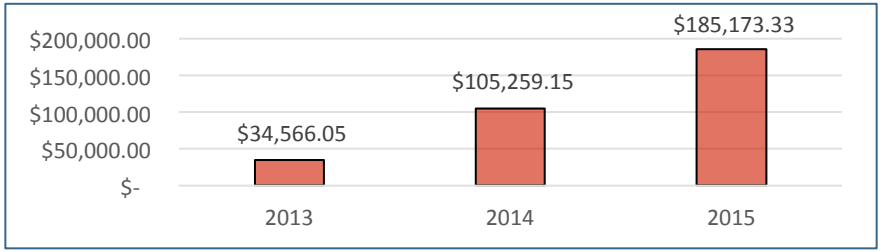
Orrstown has adopted a **cloud-first model for technology** solutions

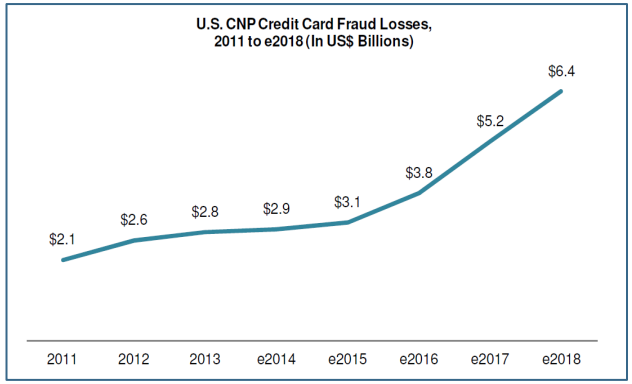
Orrstown has been running splunk>cloud for over 1 year to facilitate **operational and security data analytics**

Orrstown continues to explore opportunities to use splunk>cloud to solve **additional problems such as Fraud**

Overview of the Problem

Amount **charged off due to Card Fraud** has grown **435%** in the past 3 years





This pattern represents a **similar if not conservative experience as our peers**

Many **solutions** offered to smaller banks **cannot keep pace with the fraud** patterns

ORRSTOWN BANK

Fraudulent Behavior Patterns

Attacker steals card data from…

…and local merchants

Card data sold on the black market, cloned, and used….

Often the first fraudulent transaction is followed shortly by many other transactions

Smarter criminals are selling cards back into the local area from which the card was stolen to evade fraud detection

Fraudulent card present transactions usually occur at…

- Grocery stores
- Pharmacies
- Walmart
- GameStop
- Best Buy

Card not present transactions at…

- Apple
- On-line electronics retailers
- Travel

ORRSTOWN BANK

Gaps in Current Fraud Solutions

Processors' used by many community banks **provide only rudimentary fraud detection** capabilities

Those processors who do provide more sophisticated capabilities such as FICO Falcon are **slow to adopt full automation and all the capabilities**

Expert information about the customer is **often not considered** by the model in the risk calculations

Consortium models are generic and cannot adapt to local market conditions or local behaviors as quickly as this model.

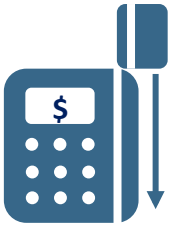
The **fully functional models are expensive** to operate

ORRSTOWN BANK

Orrstown Using Prelert and Splunk to Combat Fraud

ORRSTOWN BANK

Orrstown has a cloud-first model and as such chose Splunk's Cloud offering to consolidate operational analytics and security analytics into a single platform



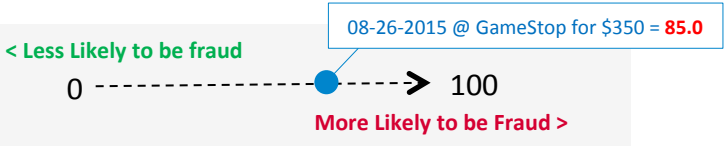
Financial Transactions

splunk>cloud

prelert

Prelert quickly calculates statistical anomalies across large datasets and provides a model that constantly learns and adapts with more information and expert intelligence.

- **Calculates anomalies across multiple dimensions including…**
  - How much was spent
  - When the transaction occurred
  - Where the transaction occurred (geographically)
  - At which merchant the transaction occurred
- **Factors in Expert Information**
  - Known Fraud Patterns
  - Confirmed Customer Travel Plans
- **Scores every transaction on the likelihood of Fraud**



Who is Prelert?

LEADER IN MACHINE LEARNING BEHAVIORAL ANALYTICS

100+ CUSTOMER INSTALLATIONS

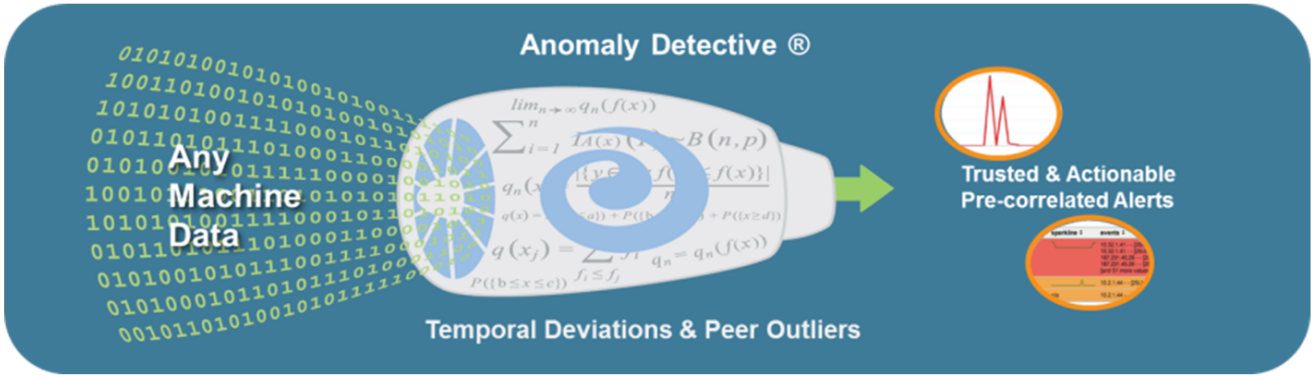
PARTNERSHIPS

Prelert Disclaimer: This presentation may include descriptions of future product capabilities. Prelert is under no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.

prelert

Prelert Anomaly Detective® for Splunk® Overview

- Automated Machine Learning Establishes "**Normal**" Baseline
- Sophisticated Statistical Modeling Identifies "**Abnormal**" Activity
- Broadly Applicable Across Source Types: Network, Server, Device, Application, User Logs



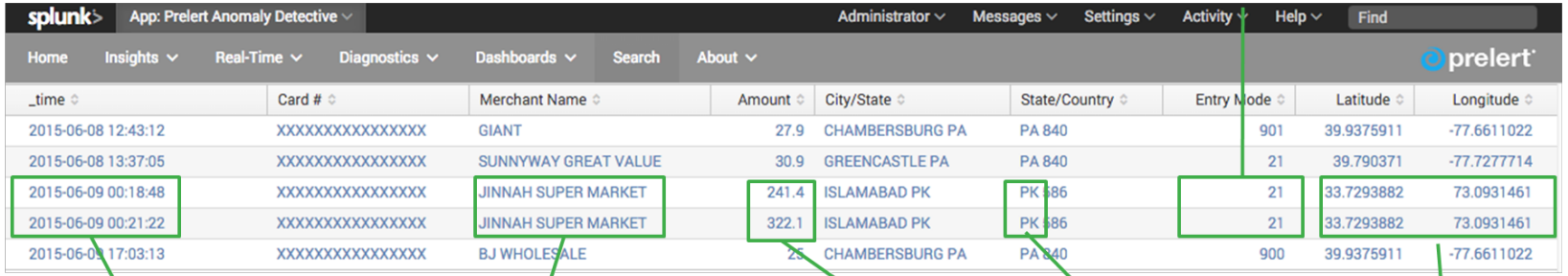
- Accurately Analyzes TB's/Day in Near Real-time
- Reduces Manual Effort, Human Error, and False Positives
- Finds Problems that Traditional Solutions Miss

prelert

Card Fraud Prediction – Technical Overview

- **Goal – Compute likelihood of a transaction begin fraudulent**
- **Card transactions consist of >100 fields**

Field values often fraud related



_time	Card #	Merchant Name	Amount	City/State	State/Country	Entry Mode	Latitude	Longitude
2015-06-08 12:43:12	XXXXXXXXXXXXXXX	GIANT	27.9	CHAMBERSBURG PA	PA 840	901	39.9375911	-77.6611022
2015-06-08 13:37:05	XXXXXXXXXXXXXXX	SUNNYWAY GREAT VALUE	30.9	GREENCASTLE PA	PA 840	21	39.790371	-77.7277714
2015-06-09 00:18:48	XXXXXXXXXXXXXXX	JINNAH SUPER MARKET	241.4	ISLAMABAD PK	PK 586	21	33.7293882	73.0931461
2015-06-09 00:21:22	XXXXXXXXXXXXXXX	JINNAH SUPER MARKET	322.1	ISLAMABAD PK	PK 586	21	33.7293882	73.0931461
2015-06-09 17:03:13	XXXXXXXXXXXXXXX	BJ WHOLESALE	25	CHAMBERSBURG PA	PA 840	900	39.9375911	-77.6611022

Unusual transaction time for card

Multiple transactions to same merchant in short period

Unusual transaction amount rate

Rare country

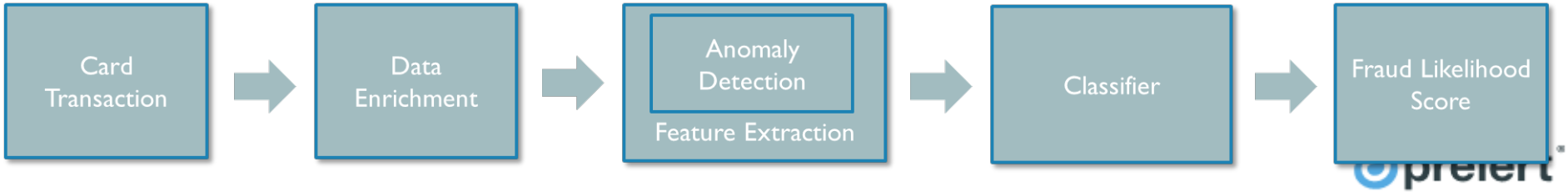
Rare lat/lng for card

- **How to calculate probability of a transaction being fraud?**

Card Fraud Prediction – Technical Overview

- **Available information:**
  - Historical transactions
  - List of transactions that were reported as fraud by card holders

- **Overall approach:**
  - Use unsupervised and supervised machine learning to model transactions and to calculate probability of a transaction being fraud – in real-time
    - Select and calculate features that differentiate fraud – e.g. rarity of location, distance between transactions, transaction pattern etc.
    - Train classifier based on historical fraud (and update as new fraud occurs)
    - Calculate the likelihood of a transaction being fraudulent based on model



Card Fraud Prediction – Feature Engineering

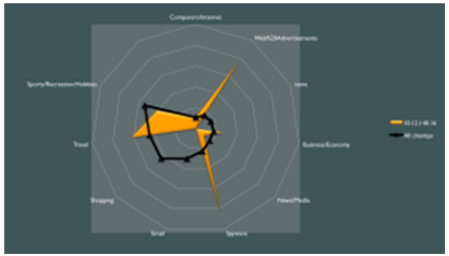
- **Key to this process is creation of *accurate* and *differentiating* features**

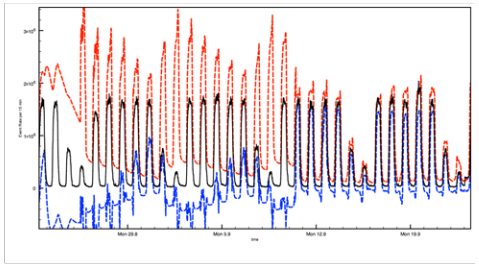
- **Inputs into this process:**
  - Patterns learned from the marked-up Fraud data and non-Fraud data
  - Expert knowledge into what is causing Fraud
  - Shared community knowledge

- **Features types:**
  - Data attributes e.g. ATINTL='F' (International Flag = 'F')
  - Derived statistical features e.g. rarity of location, probability of transaction pattern
  - Derived boolean features e.g. velocity required between transactions > 900km/h

Card Fraud Prediction – Derived Features

- **Prelert's core capability is accurate anomaly detection for machine data**
  - Unsupervised machine learning – automatically learns normal patterns in the data
    - Metrics and categorical data
    - Individual, peer and population analysis
    - Automatic fitting of multi-modal distributions, detection of periodic trends, clustering of peers
    - Scales to TB's/day, native Splunk app
- **Use anomaly detection to create robust, accurate generalized features for Prelert's classifier**
  - Unusual transactions for a card
  - Unusual transactions for all cards
  - Unusual transactions for groups (peers) of cards



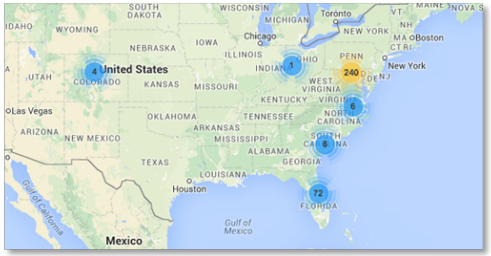


prelert

Card Fraud Prediction – Example Feature

- **Location is a good differentiator for 'Card-Present' fraud**
  - Enrich data with latitude/longitude via Google GeoAPI:
    - https://maps.googleapis.com/maps/api/geocode/json?address=BJ+WHOLESALE%2CCHAMBERSBURG+PA%2CUSA&key=…

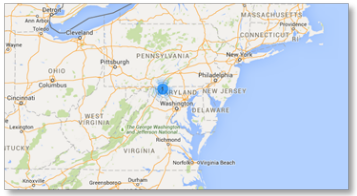
Merchant Name	City/State	State/Country	Latitude	Longitude
BJ WHOLESALE	CHAMBERSBURG PA	PA 840	39.9375911	-77.6611022



Card travelled 6478km in ~1h
1.66km/s ~ 3700mph

  - Enrich data with distance/velocity between current transaction and last 'card-present' transaction

_time	Merchant Name	City/State	State/Country	Latitude	Longitude	Distance	Velocity
2015-03-29 09:06:07	Starbucks	HAGERSTOWN MD	MD 840	39.6417629	-77.7199932	48.5578426158	0.00784836635136
2015-03-29 09:29:00	Wal-Mart Super	HAGERSTOWN MD	MD 840	39.6417629	-77.7199932	0.0	0.0
2015-03-29 10:33:46	Hotel NH Ciutat de Vic	VIC ES	ES 724	41.9304373	2.2544335	6478.32141714	1.66709249026
2015-03-29 21:33:12	Wal-Mart Super	SHIPPENSBURG PA	PA 840	40.0506453	-77.5202647	6440.6211605	0.162781710572
2015-03-30 11:07:29	McDonalds	WILLIAMSON NY	NY 840	43.2239229	-77.1861277	353.942346965	0.00724445518482



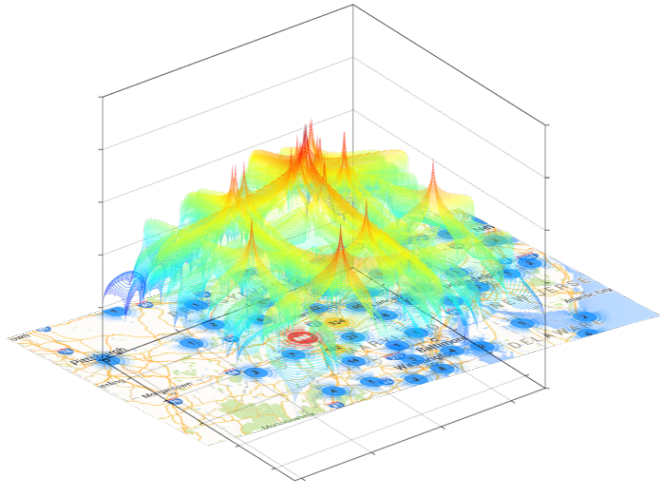
09:29:00 – Hagerstown, MD, USA

10:33:46 – Vic, Spain

prelert

Card Fraud Prediction – Location Feature

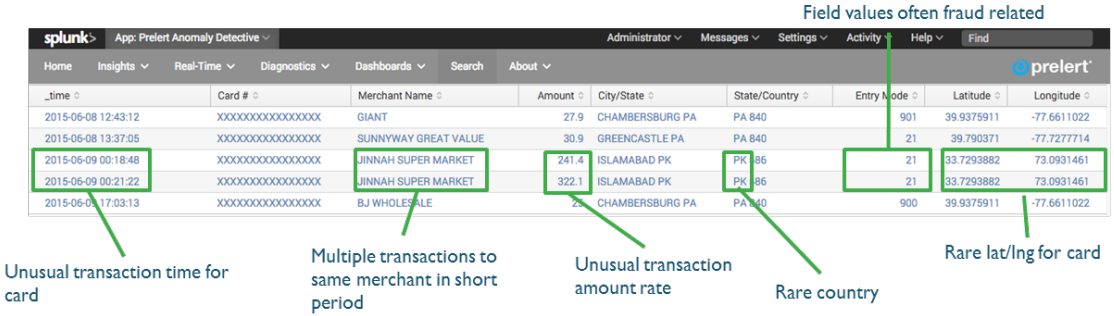
- **Once data is enriched, we need to calculate the probability of the location given the card's history and the history of all cards**
- **Automatically learn 2d multi-modal probability distributions – in real-time**
- **Useful features example:**
  - *p(location | card history) = 6e-13*
  - *p(location | all cards history) = 0.02*
  - *p(distance | card history) = 0.04*
  - *p(velocity | card history) = 2e-8*
  - *velocity > 900km/h = false*
  - *Is international transaction (field ATINTL="*") = true*



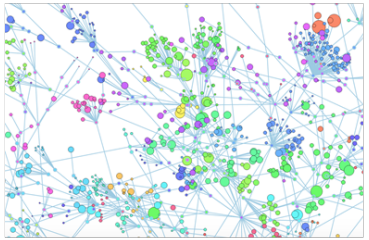
- **However, location is a useful feature, but increasingly fraud is being performing locally**
  - Use location in conjunction with features such as 'transaction patterns' to predict fraud

prelert

Card Fraud Prediction – Classification



- **Workflow:**
  - Select and compute features (including statistical features)
  - Train classifier on marked-up data
  - For new transactions
    - Compute features
    - Run features through classifier to compute risk score (0-100)
  - For new known fraud transactions
    - Retrain classifier on new mark up overnight



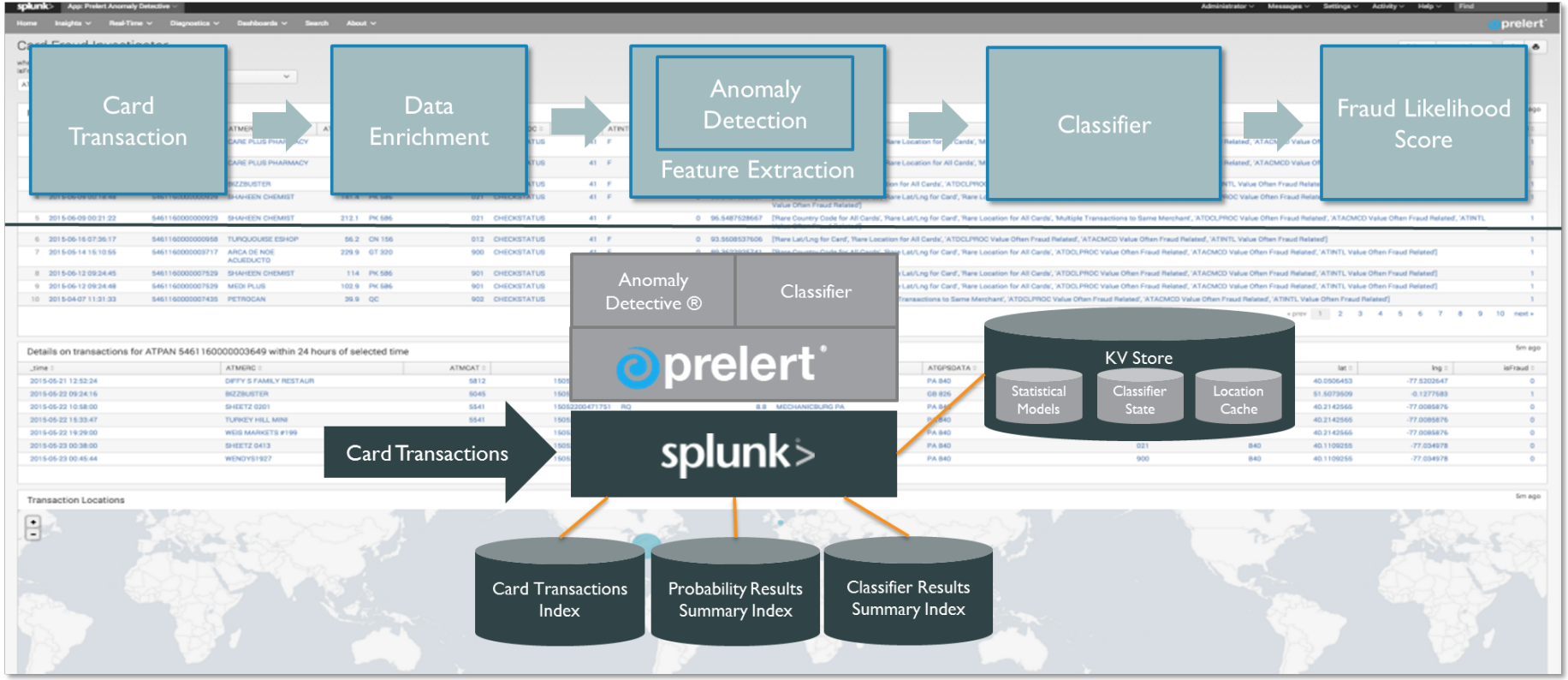
Schematic representation of trained classifier

prelert

Card Fraud Prediction – Challenges

- **Training data can be very sparse and skewed (i.e. very little fraud to non-fraud), for example in a time period there were:**
  - 422 fraud transactions
  - 1,293,790 non-fraud transactions
- **Not all fraud is reported**
- **Accuracy and quality of features is key**
  - Inaccurate anomaly detection with large numbers of false positives and false negatives has a large effect on the quality of the predictions.
    - For example, using simple approaches such as *anomaly is anything > mu + 2.5*sigma* can result in features being ignored by the classifier
- **Very easy to overfit classifier**
  - i.e. create a model that doesn't generalise and will only work well on the training dataset
- **Needs to be implemented as close to real-time as possible**
  - Fraud often follows fraud…

Card Fraud Prediction – App Architecture

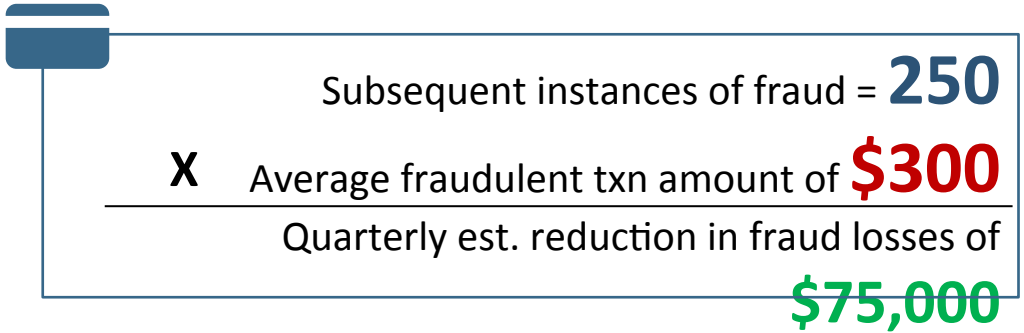


Results and the Business Case

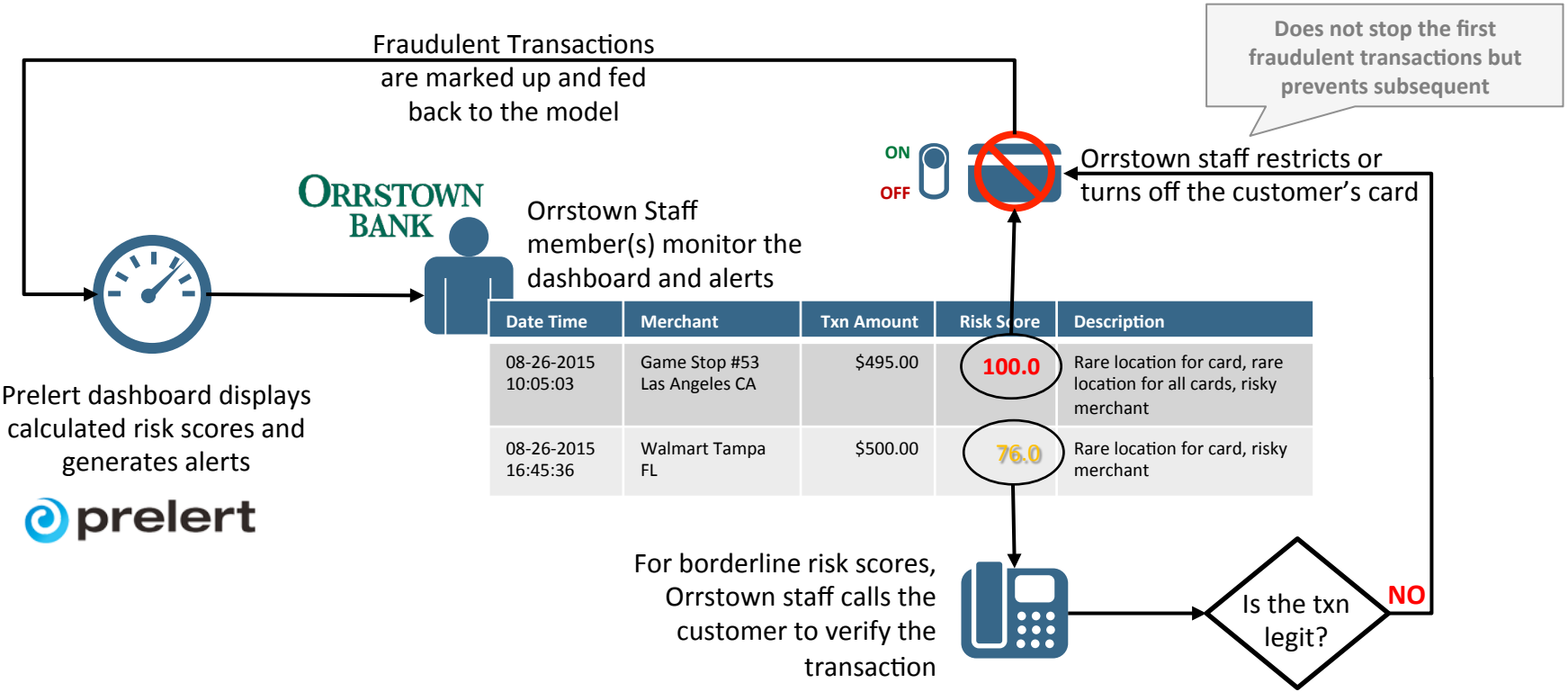
Initial experiment with **1 quarter's worth of transactions** identified…

- Approximately **50%** of the fraudulent *card present* transactions
- A small population of only **330** false positives
- Of the fraudulent transactions identified, there were **250 instances of subsequent fraudulent transactions** occurred using the same card

These are the fraudulent txns we can stop



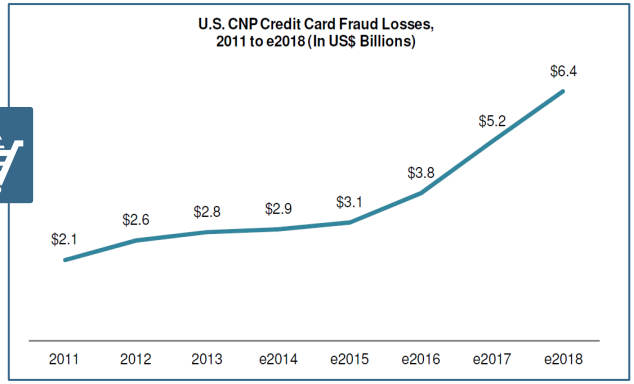
Operationalizing the Model



What's Next?

EMV Cards will become more predominant in the US and will **reduce card present fraud**

Card Not Present (CNP) Fraud expected to grow and therefore is Orrstown's next challenge



Tokenization

3-D Secure

Verified by VISA
MasterCard SecureCode
AMERICAN EXPRESS SafeKey

Techniques to **prevent CNP Fraud**

Not 100% effective

Need to grow and expand **behavior analysis with** prelert

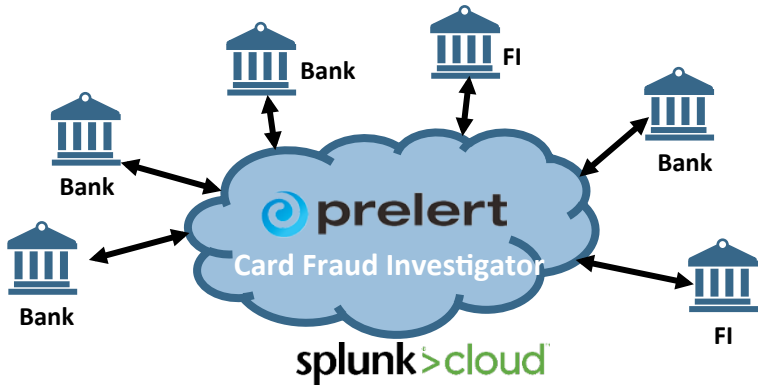
Potential

Potential for Orrstown Bank **...**

- Continue to use Prelert as an **additional layer of fraud control** beyond what our process offers
- Explore feeding **additional information** (e.g. on-line transfers, etc) into prelert via splunk>cloud **to identity other customer behavior anomalies**

Potential for the Industry...

- A **centralized Fraud Investigator service** to which many banks can subscribe
- The **model improves by** learning from a larger volume of information
- Each bank can still maintain the ability to **influence the model based on their local experiences**



THANK YOU

splunk>