Predictive System Performance Data Analysis
Application of machine learning on performance data

Jasmin Nakic, Lead Software Engineer
Samir Pilipovic, Senior Software Engineer

jnakic@salesforce.com
spilipovic@salesforce.com
Agenda
Presentation and Tutorial

Welcome

**Audience**: Sysadmins, performance engineers and developers  
**Level**: Beginner

Introduction

Introduction to Predictive Performance Analytics  
Data Visualization using Salesforce Wave

Hands-on

Prepare Input Data  
Build and Compare Predictive Models  
Generate Dynamic Alerts

Summary

Q&A
Forward-Looking Statements

Statement under the Private Securities Litigation Reform Act of 1995:

This presentation may contain forward-looking statements that involve risks, uncertainties, and assumptions. If any such uncertainties materialize or if any of the assumptions proves incorrect, the results of salesforce.com, inc. could differ materially from the results expressed or implied by the forward-looking statements we make. All statements other than statements of historical fact could be deemed forward-looking, including any projections of product or service availability, subscriber growth, earnings, revenues, or other financial items and any statements regarding strategies or plans of management for future operations, statements of belief, any statements concerning new, planned, or upgraded services or technology developments and customer contracts or use of our services.

The risks and uncertainties referred to above include – but are not limited to – risks associated with developing and delivering new functionality for our service, new products and services, our new business model, our past operating losses, possible fluctuations in our operating results and rate of growth, interruptions or delays in our Web hosting, breach of our security measures, the outcome of any litigation, risks associated with completed and any possible mergers and acquisitions, the immature market in which we operate, our relatively limited operating history, our ability to expand, retain, and motivate our employees and manage our growth, new releases of our service and successful customer deployment, our limited history reselling non-salesforce.com products, and utilization and selling to larger enterprise customers. Further information on potential factors that could affect the financial results of salesforce.com, inc. is included in our annual report on Form 10-K for the most recent fiscal year and in our quarterly report on Form 10-Q for the most recent fiscal quarter. These documents and others containing important disclosures are available on the SEC Filings section of the Investor Information section of our Web site.

Any unreleased services or features referenced in this or other presentations, press releases or public statements are not currently available and may not be delivered on time or at all. Customers who purchase our services should make the purchase decisions based upon features that are currently available. Salesforce.com, inc. assumes no obligation and does not intend to update these forward-looking statements.
<table>
<thead>
<tr>
<th>Understand</th>
<th>Predict</th>
<th>Alert</th>
<th>Visualize</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Understand</strong> short and long term trends</td>
<td><strong>Make predictions</strong> using machine learning</td>
<td><strong>Generate alerts</strong> based on dynamic thresholds</td>
<td><strong>Visualize</strong> predictions vs. real metrics</td>
</tr>
<tr>
<td><strong>Find</strong> periodic pattern in performance metrics</td>
<td><strong>Detect anomalies and exceptions</strong> on current system performance data</td>
<td><strong>Estimate impact</strong> of difference between predicted and real values</td>
<td><strong>Deliver</strong> data to other teams and executives</td>
</tr>
</tbody>
</table>
Understanding Input Data

- Data show daily and weekly periodic patterns
- Long running trend with gradual increase
- Exceptions from typical daily metric shape
## Selecting the Feature Set

Features can be generated or collected from external systems.

### Timestamp Formats
- **Timestamp as a string:**
  - `2016-05-08 19:00:00`
- **Ordinal date as decimal date plus time:**
  - `736092.791667`

### Timestamp Components
- Year, Month, Day
- Hours, Minutes, Seconds
- DayInWeek, WeekInYear
- Quarter

### Binary Features
- `isMonday`, `isTuesday`, ...
- `isSunday`
- `isHour0`, `isHour1`, `isHour2`, ..., `isHour23`
- `H_d_i_h_j` where $d_i$ is for weekday in range 0..6 and $h_j$ is for hour in range 0..23
- `isHoliday`

Use ordinal time for long term trends  
Commonly used in Analytics  
Binary features have values 0 or 1

---

A *feature* is an individual measurable property of a phenomenon being observed. (Wikipedia)
Linear Regression Models

Examples of linear regression models that apply to performance data

Simple linear regression model:

\[ y = c_0 + c_1 x \]

Model with periodic trigonometric features:

\[ y = c_0 + c_1 x + c_2 \sin(x \cdot 2 \pi / f + t) \]

Where \( t \) is time offset and \( f \) is frequency

Model with binary features:

\[ y = c_0 + c_1 x_1 + c_2 x_2 + \ldots + c_n x_n \]

Where \( x_n \) can be 0 or 1

Linear regression helps you find coefficient values \((c_0, c_1, \ldots, c_n)\)
# How Good is the Model?

Generate the “score” for each model to see how well it fits input data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training Score</th>
<th>Test Score</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Linear</td>
<td>0.0014</td>
<td>-0.0052</td>
<td>Not useful</td>
</tr>
<tr>
<td>Trigonometric</td>
<td>0.7284</td>
<td>0.7148</td>
<td>Periodic, but functions are complicated</td>
</tr>
<tr>
<td>Prediction Hour/Day</td>
<td>0.8280</td>
<td>0.8048</td>
<td>Each day has the same pattern</td>
</tr>
<tr>
<td>Prediction Hour/Week</td>
<td>0.9240</td>
<td>0.8918</td>
<td>Does not make good prediction on holidays</td>
</tr>
<tr>
<td>Prediction incl. Holidays</td>
<td>0.9520</td>
<td>0.9444</td>
<td>Does not make good prediction on Sundays</td>
</tr>
<tr>
<td>Prediction incl. Sundays</td>
<td>0.9522</td>
<td>0.9504</td>
<td>Optimal</td>
</tr>
</tbody>
</table>

- Models with more relevant features tend to have higher score
- Test data score is the measure of model quality
Visualization in Salesforce Wave

Effective visualization of predictive data

- Compare actual metrics to prediction
- Display alerts for detected anomalies
- Correlate performance metrics with business data
- Build the dashboard to share with other teams and executives
Dynamic Alerts

Anomaly detection for near real time notification

- Typically alerts are based on static threshold values, for example higher than 85%.
- If the value is higher (or lower) then predefined threshold, generate the alert.
- We want to separate data noise from exception.
- Cover false positives and false negatives if possible.
- Define the impact, for example: if a metric is for three hours more than X in absolute value and more than Y in percentage.
Machine Learning Development Process

From raw data to predictions

- Training Data
- Select Feature Set
- Build the Model
- Validate the Model
- Is Model OK?
- Yes
- Model
- Load
- Predictions
- No
- Iterative Process
- Convert input data to a format accepted by ML tools
- Measure the model score until it stops improving

A **feature** is an individual measurable property of a phenomenon being observed. (Wikipedia)
What did we cover today?

**Develop** predictive system applying **machine learning** methods

- Prepare the training data set,
- Build the predictive model,
- Predict the metric value for the test data set,
- Measure the quality of the predictive model

**Analyze** input data and predictions to solve business needs

- Find short and long term performance trends
- Implement a simple alert system
- Visualize data using Salesforce Wave
- Deliver Wave dashboard components to team members and executives
Next Steps and Resources

Get Demo Scripts from GitHub
https://github.com/sfperfdemo/vel2017-ml-wave

Get Salesforce Wave Utilities
https://github.com/forcedotcom/Analytics-Cloud-Dataset-Utils

Visit Salesforce Wave Tutorial
http://www.salesforce.com/analytics-cloud/overview/

Explore Machine Learning in Python (scikit)
http://scikit-learn.org
thank you

Family Reunion