PyConf Talk
Data Science
7-Dec-2019
AGENDA

- Introduction
- Survey
- Data Science Foundation Methodology
- Why Python
- Data Science - Use Cases
- Learnings so far
Suryanarayana Ambatipudi

- 20+ years
- Data Solutions, Implementation & Development
- Passionate about ML & AI last 6 years
- Worked in Diverse Domains like Brokerage Services, Employer Services, Supply Chain & Banking
- Hobbies: Yoga, Teaching & Reading Philosophy
What is the most important aspect in a Data Science Solution?

- Data
- Business
- Algorithms
- Just Chill
My ORDER

► BUSINESS especially the DECISION PROCESS

► DATA

► ALGORITHMS

*Chilling and Enjoying in the entire process
Barely enough data

Source - Machine Learning on Azure
Enough data

Source - Machine Learning on Azure
Data

- Relevant vs. Irrelevant Data

- Sufficient Information

<table>
<thead>
<tr>
<th>Grill temperature (F)</th>
<th>Weight of beef patty (lb)</th>
<th>Burger rating (out of 10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>550</td>
<td>.24</td>
<td>5.6</td>
</tr>
<tr>
<td>725</td>
<td>.45</td>
<td>9.4</td>
</tr>
<tr>
<td>600</td>
<td></td>
<td>8.2</td>
</tr>
</tbody>
</table>
Data Science Foundation Methodology

Source: IBM WHITE PAPER Published in March-2016
Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

Why Python

Consolidated from many sources
Why Python

- Easy to learn and use
- Swiss Army knife for the coding
- Integrates well with other Big Data tool sets
- Awesome libraries Data processing, Analysis and Deployment
- Great Community to Support !!
High LEVEL USE CASES

Predictive
- Workforce Design
- Service Recommendation Engine
- **Real Time Defect Prediction**
- Supplier Behavior

Prescriptive/AI
- Pallet Optimization
- Barcode Reader (Localization)
- Defect/Scratch Detection
- Smart Assigner (NLP)
Learnings SO FAR

- Assess Business Impact early on
- Time bound Experimentation
- Engage with Ground Zero Employee that will use these insights
- Validate Data Upfront - “Do not believe anyone “
- Focus on Data Collection if not being done already
- Emphasize this is a “journey” and “not a destination”
- Do not get demotivated if results are not in line with expectations
THANK YOU

Mail ID - Suryanarayana_Ambatipudi_2014@cba.isb.edu
Linkedin : https://www.linkedin.com/in/avssm/