A Day in the Life of a Data Scientist: How do we train our teams to get started with AI?

Francesca Lazzeri & Jaya Mathew

@frlazzeri @mathew_jaya
What is Artificial Intelligence (AI)?

Why AI?

How to get started with AI

Understanding the ML workflow

Suggested tools for AI development

AI usage in marketing

Fraud management use case
What is AI?
What is AI? – To sense, comprehend and act

**Sense**
Computer vision and audio processing, for example are able to actively perceive the world around them by acquiring and processing images, sounds and speech. The use of facial recognition at border control kiosks is one practical example of how it can improve productivity.

**Comprehend**
Natural language processing and inference engines can enable AI systems to analyse and understand the information collected. This technology is used to power the language translation feature of search engine results.

**Act**
An AI system can take action through technologies such as expert systems and inference engines or undertake actions in the physical world. Auto-pilot features and assisted-braking capabilities in cars are examples of this.

Emerging AI technologies
- Computer vision
- Audio processing
- Natural language processing
- Knowledge representation
- Machine learning
- Expert systems
- Virtual agents
- Identity analytics
- Cognitive robotics
- Speech analytics
- Recommendation systems
- Data visualization

Source: Accenture: Why artificial intelligence is the future of growth, April 2016

#StrataData
@frlazzeri @mathew_jaya
Why AI?
Three major trends are converging

Data

Cloud

Intelligence

#StrataData

@frlazzeri

@mathew_jaya
From data to decisions and actions

Descriptive (Reports)
Diagnostic (Interactive Dashboards)
Predictive (Machine Learning)
Prescriptive (Recommendations & Automation)

- What happened?
- Why did it happen?
- What will happen?
- What should I do?

Insight

#StrataData  @frlazzeri  @mathew_jaya
Digital transformation is driving new business value

- Engage your customers
- Empower your employees
- Transform your products
- Optimize your operations

Systems of Intelligence

@frlazzeri  @mathew_jaya
How to get started with AI
I want to use AI – How can I get started?

aka.ms/AlCognitiveServices
aka.ms/AlCustomModels
Being Obsessed with Data

<table>
<thead>
<tr>
<th>Question</th>
<th>Data measures what you care about</th>
<th>Data is accurate</th>
<th>Data is connected</th>
<th>A lot of data</th>
</tr>
</thead>
<tbody>
<tr>
<td>E.g. Predict whether component X will fail in the next Y days</td>
<td>E.g. Identifiers at the level you are predicting, relevant data collected &amp; feature engineering using domain knowledge</td>
<td>E.g. Failures are really failures, human labels on root causes</td>
<td>E.g. Machine information linkable to usage information</td>
<td>E.g. Will be difficult to predict failure accurately with few examples</td>
</tr>
</tbody>
</table>

#StrataData  
@frlazzeri  
@mathew_jaya
### Business scenario | Key decision | Data Science question
--- | --- | ---
Energy forecasting | Should I buy or sell energy contracts? | What will be the long/short-term demand for energy in a region? 
Customer churn | Which customers should I prioritize to reduce churn? | What is probability of churn within X days for each customer? 
Personalized marketing | What product should I offer first? | What is the probability that customer will purchase each product? 
Product feedback | Which service/product needs attention? | What is social media sentiment for each service/product?
Defining Performance Metrics

1. Establish a Qualitative Objective
2. Translate into Quantifiable Metric
3. Quantify the Metric Value Improvement useful for customer scenario
4. Establish a Baseline
5. Establish how to measure the improvement in the Data Science Metric
Understanding the ML workflow
Sample ML workflow

1. Understand Business Goals
2. Discover & Gather Data
3. Ingest Data
4. Understand Data
5. Transform Data
6. Create Model
7. Deploy Model
8. Monitor & Maintain Model
9. Collaboration & Version Control
10. Documentation

Build a model
- Science

Publish a model
- Operations

Consume a model

Share Results with Business Owners

Debug, Fix, Enhance, etc.

Respond to changes/lessons

Apps
Services
Data Engines

#StrataData
@frlazzeri
@mathew_jaya
Team Data Science Process
aka.ms/TeamDataScience

Data Sources ➤ Ingest ➤ Prepare ➤ Analyze ➤ Publish ➤ Consume

- Domain expert
- Solution Architect
- Data Engineer
- Data Scientist
- Visualization Expert
- Integration Engineer
- Project Manager
- Executive Sponsorship – IT & Business

#StrataData
@frlazzeri
@mathew_jaya
Suggested tools for AI development
Example Tools / Dev Environments

- **Azure Machine Learning Studio**: Serverless collaborative drag-and-drop tool for graphical machine learning development

- **Azure Machine Learning Services**: Visual AI powered data wrangling, experimentation, and lifecycle management

- **Visual Studio Code Tools for AI**: Build, debug, test, and deploy AI with Visual Studio Code on Windows and Mac

- **Azure Notebooks**: Organize your datasets and Jupyter Notebooks in a centralized library for Data Science and Analysis

- **Deep Learning Virtual Machine**: A pre-configured environment for deep learning using GPU instances
AI Solution Templates
aka.ms/AzureAIGallery

#StrataData  @frlazzeri  @mathew_jaya
Microsoft Learning
aka.ms/MicrosoftAIRepairing
AI usage in marketing
How can AI help in Marketing & Personalization?

**Input Data**
- In store
- Online activity
- Social media
- Past campaign performance

**ML & AI**
- Learn customer omni-channel preference
- Personalize web/email experience
- Preemptive chatbots to answer customer queries
- Monitor for cart abandonment, churn, retention

**Insight:**
- Personalized content for customers
- Identify high value customers

**Intelligent Action**
- Personalized website experience
- Adaptive product pricing, offers for cross/upsell
- Premium loyalty programs and service experience
- Predictive customer service via social media, chatbots

#StrataData  
@frlazzeri  
@mathew_jaya
Let us map the journey based on the stages in a customer life cycle as follows:

**Awareness; Acquisition; Engagement; Support; Monetization; Fraud/Risk management**

**Marketer’s AI Journey Map**

- **Awareness**
  - Lead scoring
  - Market mix optimization
- **Acquisition**
  - Opportunity Scoring
  - Fraud & Risk
  - Personalized offers
- **Engagement**
  - User segmentation
  - Social media analytics
  - Churn Identification & prevention
- **Support**
  - Customer Encounters problems
  - Customer Support (bot vs human)
- **Monetization**
  - Customer lifetime Value estimate
  - Recommendations

---

#StrataData

@frlazzeri

@mathew_jaya
Marketing cycle – Awareness & Acquisition

**Awareness**
- Marketing mix optimization for omni-channel budget optimization
- ML based lead-scoring models

**Acquisition**
- Opportunity scoring can help target users who are most likely to make a purchase
- AI powered content creation

#StrataData
@frlazzeri @mathew_jaya
Marketing cycle – Engagement & Monetization

- ✓ Customer segmentation based on their browsing/purchase patterns
- ✓ ML based models can help determine customers who are most likely to churn in the near future based on their behavior
- ✓ AI powered personalized recommendations
- ✓ Using all the information available on customer’s buying, browsing patterns determine the Lifetime value (LTV)
Automated answers based on a knowledge base can help decrease call center costs and predict staffing needs.

Support ticket clustering can help the team find solutions for customer complaints quickly and even report similar problems back to the product teams.

While acquiring customers, external customer data can be used to determine customers who are likely to be risky for the business in the long run.

ML & AI models can be built to predict if a customer is likely to end up in bankruptcy, being delinquency which results in non-payment for the products/services.
Fraud management use case
Credit Card Fraud Detection Data Set

- Dataset used:
  - contains transactions made by credit cards in September 2013 by European cardholders
  - presents transactions that occurred in two days, where we have 492 frauds out of 284,807 transactions

- Features $V_1, V_2, \ldots, V_{28}$: are the principal components obtained with PCA, the only features which have not been transformed with PCA are 'Time' and 'Amount'

- Feature $Time$: contains the seconds elapsed between each transaction and the first transaction in the dataset

- Feature $Amount$: is the transaction Amount

- Feature $Class$: is the response variable and it takes value 1 in case of fraud and 0 otherwise

Data Source: https://www.kaggle.com/mlg-ulb/creditcardfraud

@frlazzeri @mathew_jaya
The sample code was tested and run using the Jupyter notebook environment on a remote Azure VM (Standard F8s (8 vcpus, 16 GB memory))

The sample code is available at the following GitHub location: 
https://github.com/jayamathew/Codebase/tree/master/conferences

The outline of the code is as follows:

- Environment setup
- Import and check the dataset
- Modeling
- Save the model
- Model Operationalization

#StrataData @frlazzeri @mathew_jaya
Demo

- Import the necessary libraries and provide credentials to access the data

```python
# Import necessary components
import os
import shutil
import json

# Import re
import pandas as pd
import numpy as np
import sklearn
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.layers import Dense, Dropout, LSTM, Activation
from keras.callbacks import ModelCheckpoint, TensorBoard
from keras import regularizers
```
• Import the dataset and check the distributions of the variables
Demo

- Build any model and tune hyper parameters (if needed)
Environment setup

Import and check the dataset

Modeling

Save the model

Model Operationalization

• Save the best model for operationalization
Environment setup

Import and check the dataset

Modeling

Save the model

Model Operationalization

• Create the necessary functions for model operationalization using any tool of choice

Deployment

There are multiple options to operationalize a model, this is entirely dependent on the tools used.

Once the assets (model, schema file, scoring script etc.) are stored, we can download them into a deployment compute context for operationalization on an Azure web service. For this scenario, we will deploy this on our local context. We demonstrate how to setup this web service through a CLI window opened in the AML.

Create a model management endpoint

Create a model management under your account. We will call this automodelmanagement. The remaining defaults are acceptable. az ml account

Create a model management endpoint

az ml modelmanagement create --location <ACCOUNT_REGION> --resource-group <RESOURCE_GROUP> --name automodelmanagement

You can find the subscription name or subscription ID through the (https://portal.azure.com) under the resource group you’d like to use.

Check environment settings

Show what environment is currently active: az ml env show

If nothing is set, we setup the environment with the existing model management context first: az ml env setup --location <ACCOUNT_REGION> --resource-group <RESOURCE_GROUP> --name automodelmanagement

Then set the current environment: az ml env set --resource-group <RESOURCE_GROUP> --cluster-name automodelmanagement

Check that the environment is now set: az ml env show
Links to get started with AI

- Cognitive Services: https://aka.ms/AICognitiveServices
- Azure ML: https://aka.ms/AICustomModels
- Azure Machine Learning Studio: https://aka.ms/AzureStudio
- Azure Machine Learning Services: https://aka.ms/AMLServices
- Azure Notebooks: https://aka.ms/AzureJNotebooks
- Preconfigured Virtual Machines: https://aka.ms/AzureVirtualMachines
- Deep Learning Virtual Machine: https://aka.ms/AzureDSVM
- Team Data Science Process: https://aka.ms/TeamDataScience
- Data Source for demo: https://www.kaggle.com/mlg-mlb/creditcardfraud
- GitHub location for demo: https://github.com/jayamathew/Codebase/tree/master/conferences
Thank You!

Francesca Lazzeri & Jaya Mathew

@frlazzeri @mathew_jaya