

F S V 3 0 9

Market Prediction Using ML

Experiment with Amazon SageMaker and the Deutsche Börse Public Dataset

Binoy Das
Partner Solutions Architect
Amazon Web Services

Jed Sundwall
Manager, Open Data Program
Amazon Web Services

AWS re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Our Mission

Put machine learning
in the hands of every
developer,
data scientist and
architect.



Agenda

- Overview of Amazon SageMaker
- Overview of Deutsche Börse Public Dataset
- Hands On – Data preparation
- Hands On – Data analysis
- Hands On – Custom container on ECR to implement RNN model
- Hands On – Use DeepAR on Amazon SageMaker

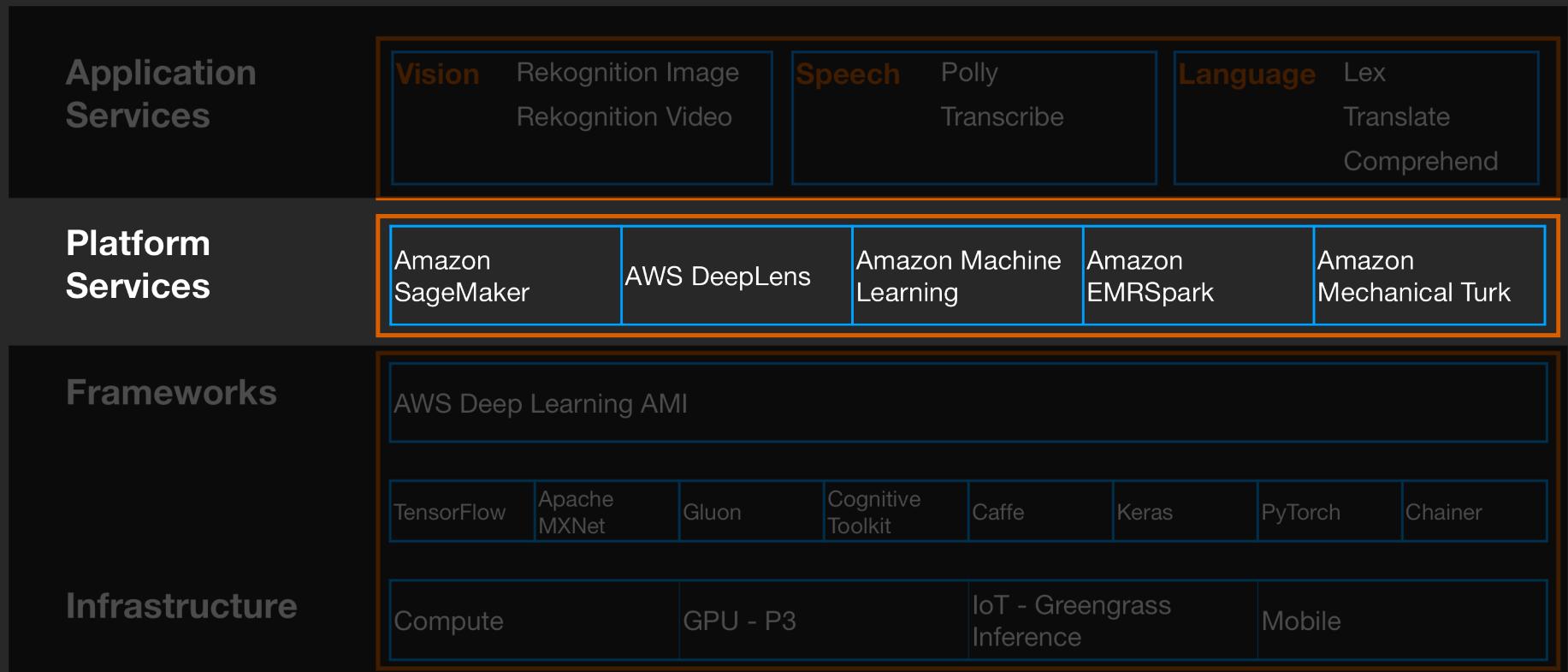
Amazon SageMaker

AWS
re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



The AWS Machine Learning Stack





Amazon SageMaker

1



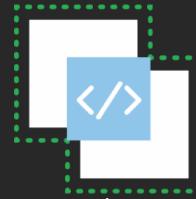
Notebook Instances

2



Algorithms

3



ML Training Service

4



ML Hosting Service

Amazon SageMaker – Value Proposition

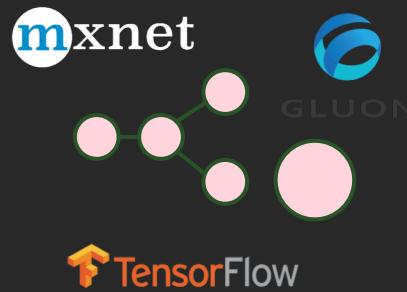
The quickest and easiest way to get ML models from idea to production



End-to-End Machine Learning Platform



Zero setup



Flexible Model Training



Pay by the second

Deutsche Börse Public Dataset

AWS
re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Stock Market data maintained by Deutsche Börse

Published on Registry of Open Data on AWS

- Available free, hosted on Amazon S3 bucket
- Trade data from Eurex and Xtera trading systems
- Continuous minute by minute update
- Opening, closing, highest and lowest price, and trading volume for every security traded
- Non-commercial license, great for learning and development

<https://registry.opendata.aws/deutsche-boerse-pds/>

Deutsche Börse Public Dataset

Source:

<s3://deutsche-boerse-xetra-pds> (for Xetra data)

<s3://deutsche-boerse-eurex-pds> (for Eurex data)

Data Dictionary:

Field	Description	Type
Mnemonic	Stock exchange ticker symbol	Text
Date	Date of trading period	Date
Time	Minute of trading to which this entry relates	Time
StartPrice	Trading price at the start of period	Decimal
MaxPrice	Maximum price over the period	Decimal
MinPrice	Minimum price over the period	Decimal
EndPrice	Trading price at the end of the period	Decimal
TradedVolume	Total value traded	Decimal
NumberOfTrades	Number of distinct trades during the period	Integer

Hands On Data Preparation

AWS
re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Begin with a Notebook

<https://github.com/aws-samples/db-reinvent-workshop>

- 1st step: Create lifecycle configurations

- Clone from GitHub
- Create folder structures
- Obtain cleaned up data

Hands On Data Analysis

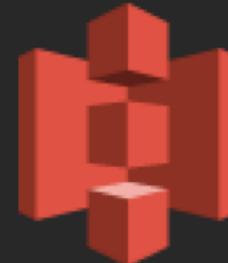
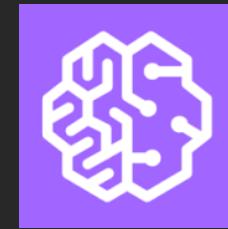
AWS
re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Explore Stock Behavior

- Visualize
 - Generate plots
 - Spot related metrics
 - Spot related stocks
- Clustering Algorithm
 - Run custom clustering algorithm
 - Find correlation
- Group Clustered Stocks
 - Retrieve metrics for clustered stocks
 - Create separate data files for clusters

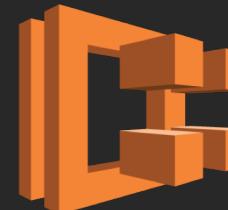
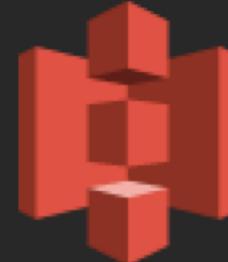


Hands On

Custom container on ECR to implement RNN Model

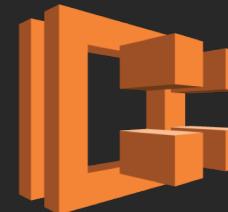
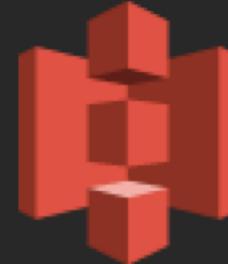
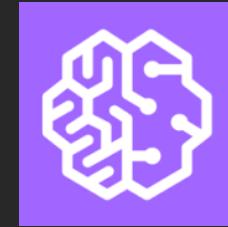
Containerize Custom MLP Code

- Data Channels
 - Peek into data
 - Load to SageMaker S3 Bucket
- Container Image
 - Bundle code and libraries
 - Create Amazon ECR Repository
- Model Training
 - Create SageMaker Estimator
 - Fit Estimator to the data



Deploy Custom MLP Model

- Model Deployment
 - Create SageMaker Predictor
 - Deploy to SageMaker Endpoint
- Inference Generation
 - Generate Prediction with test data
 - Visualize Prediction against Observed
- Performance Evaluation
 - Measure accuracy
 - Scope for improvement



Hands On

Use DeepAR on Amazon SageMaker

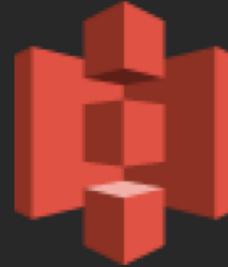
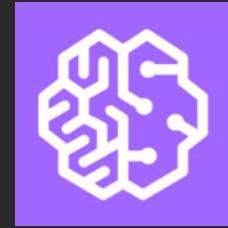
AWS
re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.



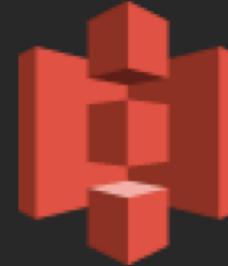
Train DeepAR Model

- DeepAR Data Set
 - Format clustered data for DeepAR
 - Create Training and Test data
- Model Training
 - Create SageMaker Training Job
 - Observe Training Behavior



Deploy DeepAR Model

- Model Deployment
 - Create Endpoint Configuration
 - Create SageMaker Endpoint
- Inference Generation
 - Generate Prediction with test data
 - Visualize Prediction against Observed
- Performance Evaluation
 - Measure accuracy
 - Repeat with different cluster



Thank you!

Binoy Das

binoyd@amazon.com

Jed Sundwall

jed@amazon.de

AWS re:Invent

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.

