INSPARI
a valantic company

Microsoft Fabric Festival

ML Engineering i Microsoft Fabric

INSPARI
a valantic company

Anders Meged
Data Engineer

Agenda

1. Today’s Data Scientist
2. Solution Scenarios – Live Demos
3. Wrap up and benefits for Data Scientists

INSPARI
a valantic company

Personal Data

Age:
30

Job Title:
Data Scientist

Company:
EnergyCorp

Kim

Data Scientist

Taking data to the next level

Personal profile

Kim is a Data Scientist at EnergyCorp who works with:

- Building advanced analytic models
- Leveraging machine learning techniques to extract valuable insights from complex dataset.

While her background includes experience in traditional data analytics, Kim now focuses on pushing the boundaries of data science, exploring cutting-edge methodologies and algorithms to uncover patterns and trends that contribute to the overall success of EnergyCorp. She enjoys diving into the intricacies of data, applying statistical rigor, and deploying predictive models to solve challenging business problems.

Skills

Python

SQL

Back-end Developer

Needs

- Manipulate data ad-hoc
- Data exploration
- Train and Tracking ML models
- Dynamic and modifiable data foundation

Pains

- Local environments – Lack of access across teams.
- Data cleaning and preprocessing
- Comparing Models
- Scaling and productionalizing
- Highlighting the benefits that models can provide.

INSPARI
a valantic company

Solution Scenarios

Starting point: Cleaning data for ML model

1. Clean data for Model training
Using data wrangler to clean data.

3. Comparing models
Comparing results of the different models.

5. Visualizing
Visualizing the predictions of the model.

2. Training models
Training Ml models.

4. Productionalizing models
Make the model work on new data.

INSPARI
a valantic company

Scenario 1: Prepare Data for Model Training

Problem

- Kim is tasked with a cleaning data for training ML model on predicting energy production.
- The Data Scientist in the team uses different libraries and techniques for cleaning.

Solution

- Kim builds a notebook report directly on the Fabric platform in the browser.
- Fabric notebook’s "Data Wrangler" functionality combines coding and no code tools.
- This solution lowers the entry level for unifying a framework for cleaning preprocessing.

INSPARI
a valantic company

Solution Scenarios

Training ML models

1. Clean data for Model training

Using data wrangler to clean data.

3. Comparing models

Comparing results of the different models.

5. Visualizing

Visualizing the predictions of the model.

2. Training models

Training Ml models.

4. Productionalizing models

Make the model work on new data.

Scenario 2: Training ML Models

INSPARI
a valantic company

Problem

- Kim needs to train ML models and keep track of the different Models.
- Kim’s colleagues also benefits from seeing the models she trains, as they can take advantage of the knowledge.
- Currently it is difficult to share results as they don’t have a platform for it.

Solution

• Kim creates an experiment to keep track of the different model which is trained.

• Fabric enables centralisation of the model training, which solves Kim's and the team's problem.

INSPARI
a valantic company

Solution Scenarios

Comparing Ml models

1. Clean data for Model training

Using data wrangler to clean data.

3. Comparing models

Comparing results of the different models.

5. Visualizing

Visualizing the predictions of the model.

2. Training models

Training ML models.

4. Productionalizing models

Make the model work on new data.

Scenario 3: Comparing Models

INSPARI
a valantic company

Problem

- Kim would usually log the training of the ML models in a file to compare different configurations of the model. But it is time consuming as it is hard to keep track of which model performs the best.

Solution

- Kim utilizes the list model to compare runs. Which is functionality of Fabric’s experiment artifact. Kim is now able to keep track of the different runs and easily create or find models created from the runs.

INSPARI
a valantic company

Solution Scenarios

Productionalizing ML models

1. Clean data for Model training

Using data wrangler to clean data.

3. Comparing models

Comparing results of the different models.

5. Visualizing

Visualizing the predictions of the model.

2. Training models

Training ML models.

4. Productionalizing models

Make the model work on new data.

Scenario 4: Productionalizing Models

INSPARI
a valantic company

Problem

- The Model is performing well enough for displaying it to management.
- Kim is tasked with applying the model to weather forecast data and making it accessible for visualization.

Solution

- Kim creates a notebook with the model's API and predicts on the forecast data and loads the data to a table in a Lakehouse on the OneLake.

INSPARI
a valantic company

Solution Scenarios

Vizualising ML models

1. Clean data for Model training

Using data wrangler to clean data.

3. Comparing models

Comparing results of the different models.

5. Visualizing

Visualizing the predictions of the model.

2. Training models

Training ML models.

4. Productionalizing models

Make the model work on new data.

Scenario 5: Visualize Data

INSPARI
a valantic company

Problem

- Kim needs to visualize the how the model can be used for management. The management is not technical profiles.

Solution

- Kim creates a simple Power BI report where she visualizes the predictions on future energy production.

INSPARI
a valantic company

Solution Wrap Up

Starting point: ML Model delivered by Data Scientist

1. Clean data for Model training

Using data wrangler to clean data.

3. Comparing models

Comparing results of the different models.

5. Visualizing

Visualizing the predictions of the model.

2. Training models

Training ML models.

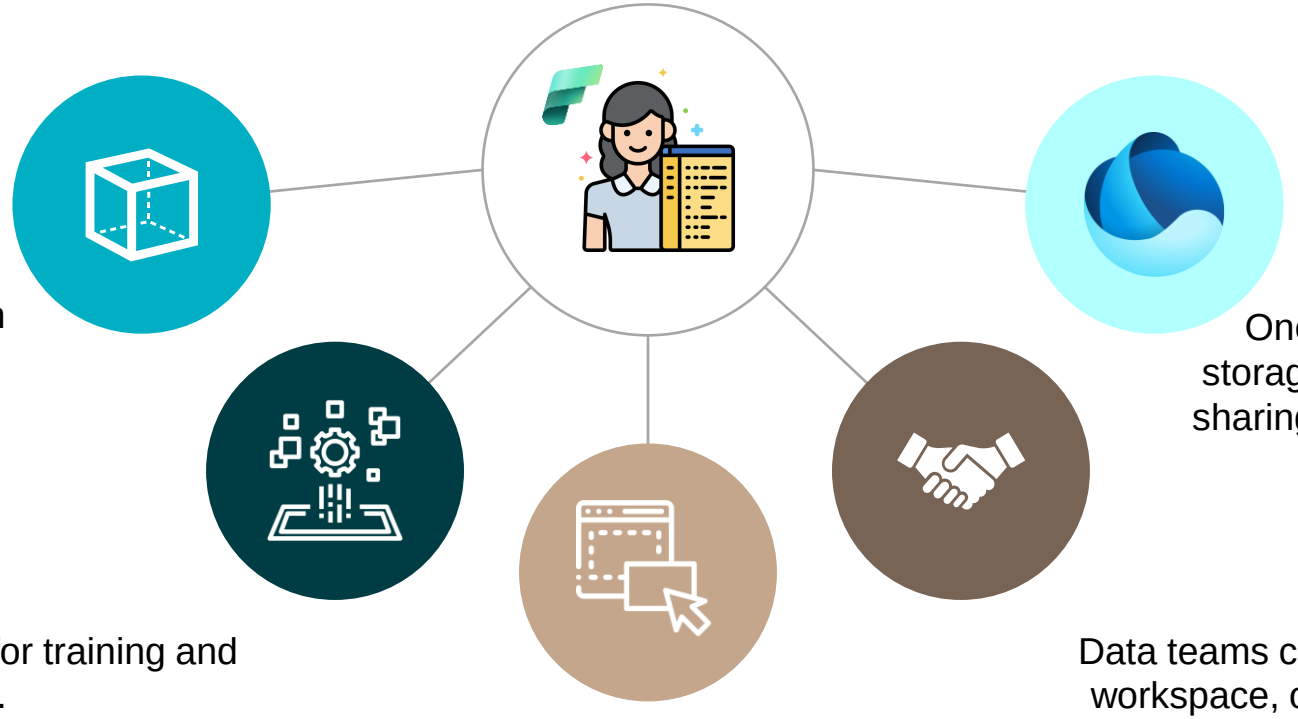
4. Productionalizing models

Make the model work on new data.

INSPARI
a valantic company

Generate value from ML models

Microsoft Fabric benefits for the Data Scientist



Data Wrangler
Easy to use data transformation and cleaning.

OneLake: OneDrive for data
OneLake provides a single, unified storage system, where discovery and sharing of data of any format between users and applications is easy.

End-to-end solution
A complete ML-platform for training and productionalizing models.

Collaboration
Data teams can collaborate in a single workspace, on the same environment settings.

Easy to Use SaaS solution
User Friendly interface with possibilities for code and no/low-code tools.