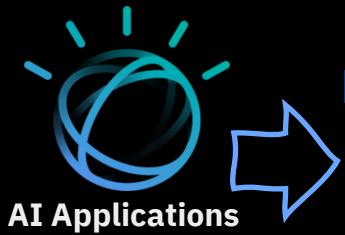


# Analytics



# The AI Ladder

A prescriptive approach to accelerating the journey to AI



**INFUSE** – Operationalize AI with trust and transparency

AI Applications

**ANALYZE** - Scale insights with AI everywhere

**ORGANIZE** - Create a trusted analytics foundation

**COLLECT** - Make data simple and accessible



Talent &  
Skills



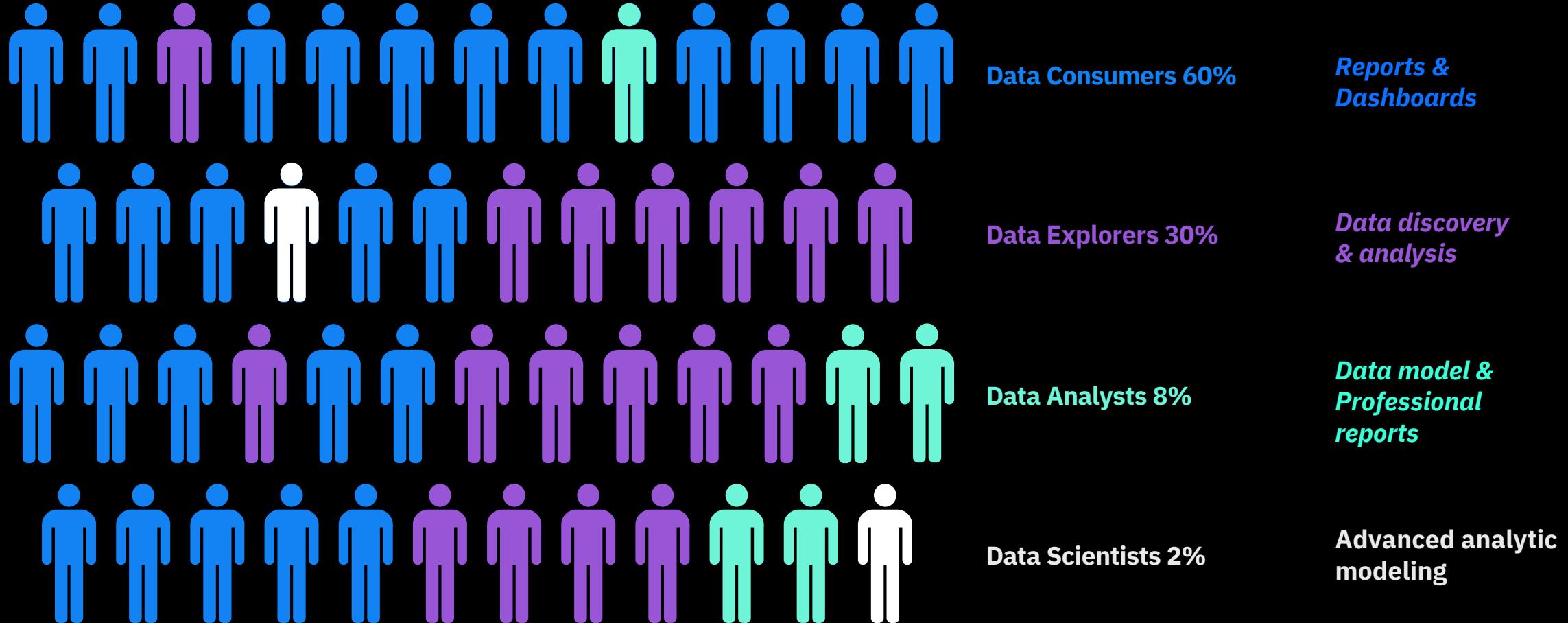
**One Platform,  
Any Cloud**

# *Cognos Dashboards*

*If I only meet your  
needs today, I am  
already behind.*

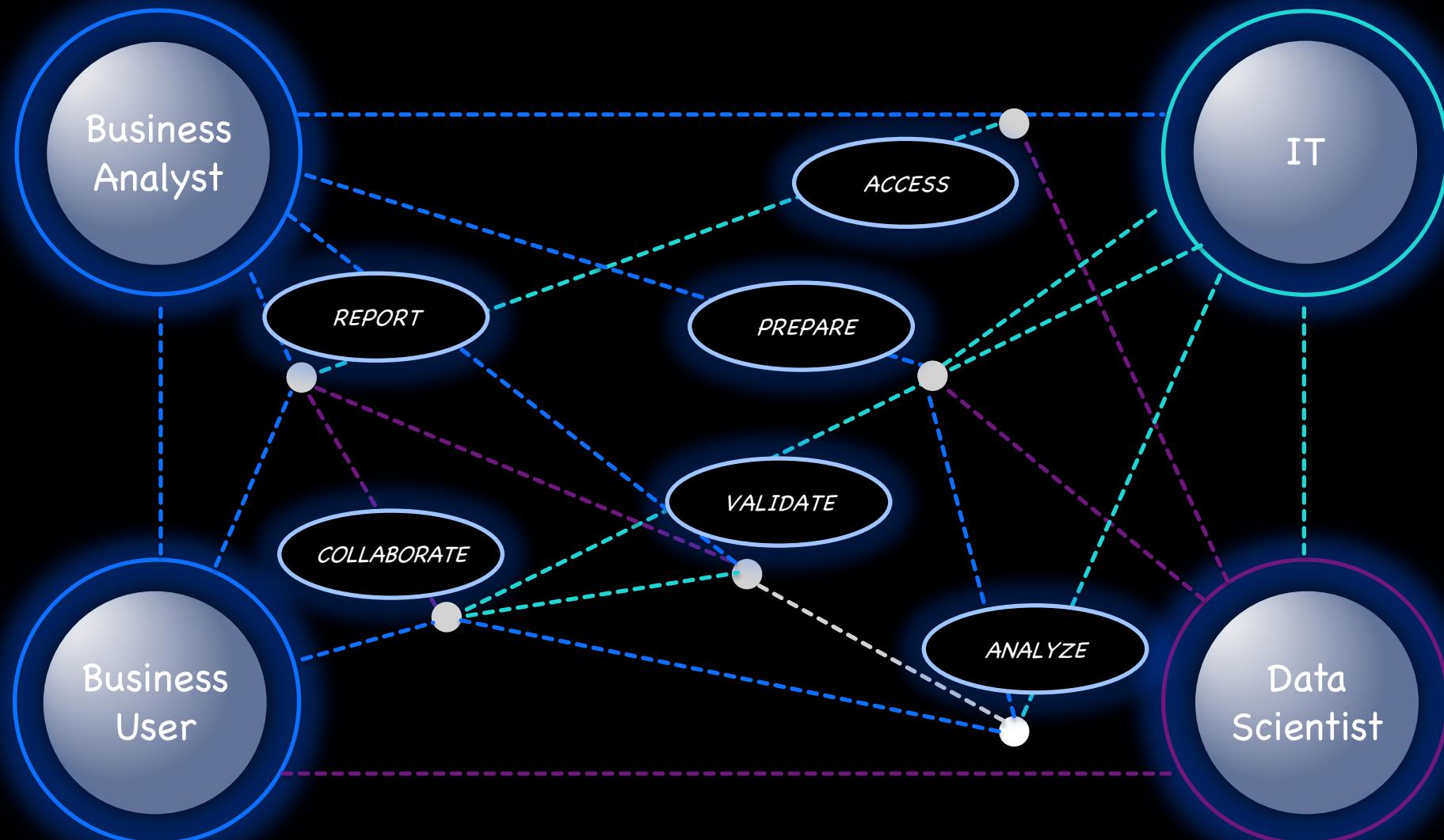
# The population of professionals is diverse

## It has experienced many changes



# Analytics Projects are Complex

Multiple steps and Multiple people



*It's Rarely a Straightforward Process*

*All the answers  
you need, in  
one place.*

# A long rich history of innovation

Impromptu  
Powerplay  
ReportNet



Cognos 8



Cognos 10



***Cognos 11.1  
AI infused***

Build on a proven platform

Metadata Modeling

Content Administration

Alerts & Notifications

Search

Reporting & Dashboarding

Office Integration

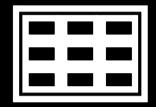
Bursting & Scheduling

Collaboration

Governance and Security

# Transform Decision Making with AI

## Enable the individual, department, organization



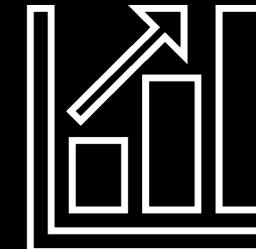
Collect  
Organize



*100's  
of questions  
every day*



*10's  
of insightful  
discoveries*



Actionable  
Insights



*10's  
of decisions  
influenced*



*100's  
of measured  
outcomes*

\$

Business Impact : Analytic Governance

\$\$\$\$

# Analytics the way you think?



Collect &  
Organize



100's  
of questions  
every day



10's  
of insightful  
discoveries



Actionable  
Insights



10's  
of decisions  
influenced



100's  
of measured  
outcomes

## Automated Data Preparation

Natural language search  
Auto-modeling  
In-tool data preparation

## Data Exploration

AI Assistant  
Advanced pattern detection  
Predictive capabilities  
AI Detected Insights  
Times Series Forecasting

## Dashboards

Automated visualizations  
Natural language generation  
Open Source Visualizations  
Storytelling  
Collaborate through Slack

## Going Beyond BI

Jupyter Notebooks & Python  
Open Source Visualizations  
IBM Planning Analytics  
IBM Decision Optimization  
Watson Studio & Watson M/L  
Watson Assistant

## Managed Reporting

Guided layouts  
Reuse existing components & styles

# Why Cognos Analytics...*Results Driven with AI*

Prepare

AI-assisted and automated to easily cleanse and combine your data in minutes

Analyze

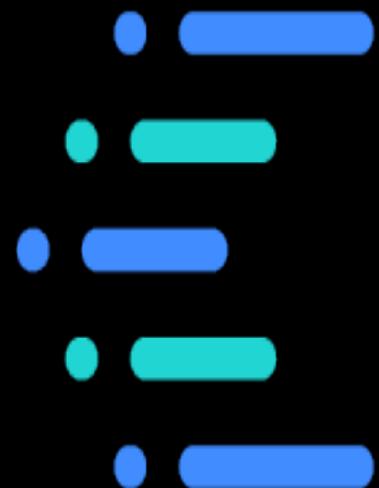
Guided exploration infused with augmented intelligence and machine learning to uncover hidden patterns and insights

Create & Share

Interactive dashboards, storytelling and pixel-perfect reporting to collaborate and share across your organization

INFUSE

Technology integration that allows you to infuse analytics into your business process



IBM  
Cognos  
Analytics

*Watson OpenScale*

# **Current Model Risk Management practices work for existing statistical or rule-based models**

**Processes rely on regular interaction between validator and model developer**

- Easy to state and translate the regulatory guidelines into requirements
- Validators can reproduce the results relatively easily

**Current systems focus more on the documentation and governance aspects of model validation**

- Who developed the model?
- What data was used to test it?
- How did data and model evolve?
- Who approved it?
- Were approved techniques used?

**Traditional statistical models are deterministic in nature or simpler to interpret and explain**

- Business rules
- Descriptive statistics
- Standard financial formulae
- Excel spreadsheets
- Linear regressions
- Decision trees

# **Validation of AI models need to augment governance with proactive testing before and in production**

## **Skills gap in order to test and validate ML/DL models**

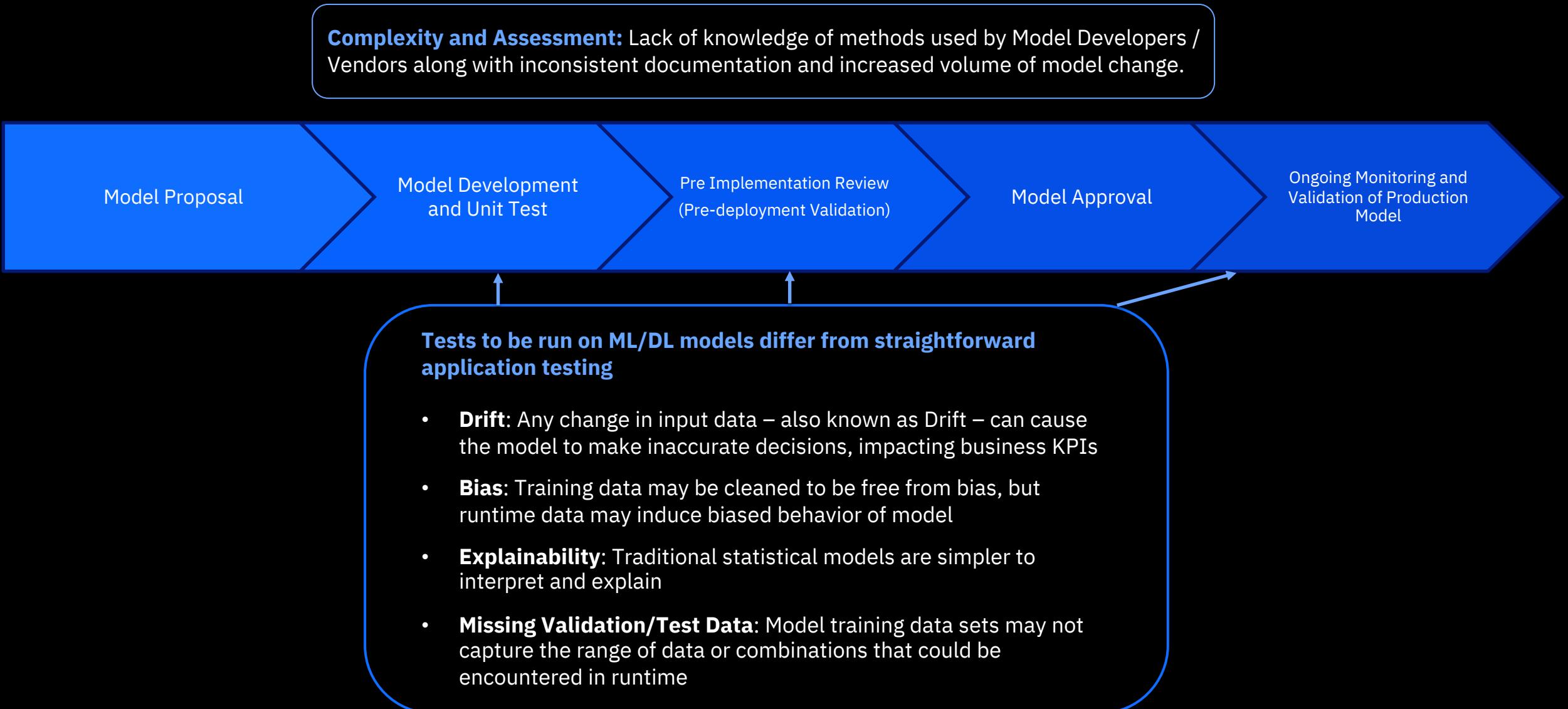
- Probabilistic nature of ML/DL models do not provide a straightforward result for interpretation
- Interpreting results and explaining to business managers and risk professionals require substantial time and effort from data scientists
- Model metrics do not convey business KPI impacts

## **Tests to be run on ML/DL models differ from straightforward application testing**

- Model training data sets may not capture the range of data or combinations that could be encountered in runtime
- Any change in input data – also known as Drift – can cause the model to make inaccurate decisions, impacting business KPIs
- Training data may be cleaned to be free from bias but runtime data may induce biased behavior of model
- Difficult to run what-if scenarios on probabilistic models without additional data sets

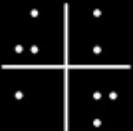
# Model Development Cycle

## *Challenges faced with ML/DL Models*



# Watson OpenScale

Validate and monitor AI models, deployed anywhere, to help comply with regulations, address internal safeguards, and mitigate business risk

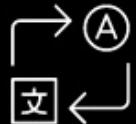


## Production monitoring for compliance and safeguards

Mitigate biased model behavior

Explain model decisions

Validate and control risk



## Ensure that models are resilient to changing situations

Detect drift during runtime

Generate specific model retraining inputs



## Align model performance with business outcomes

Actionable metrics and alerts

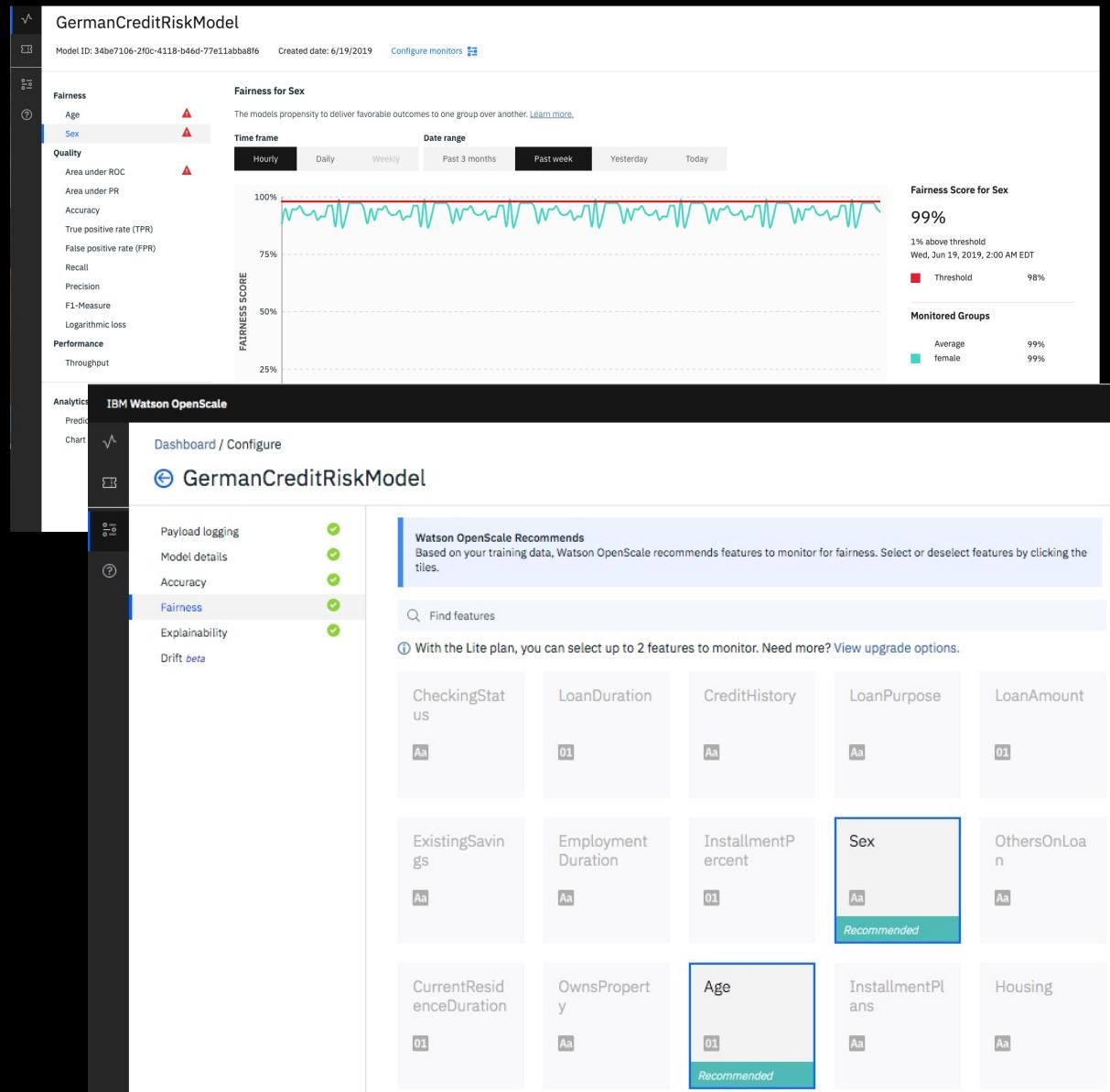
# Bias Detection

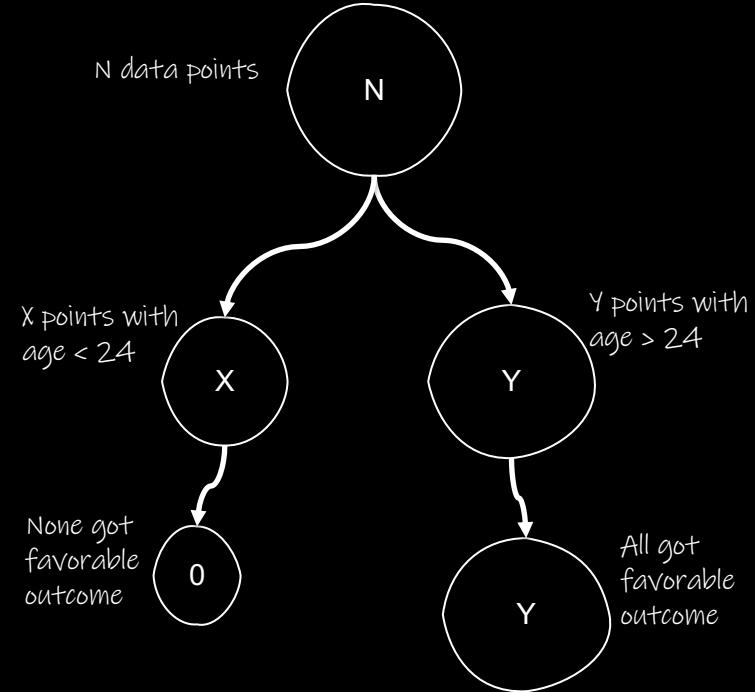
OpenScale enables enterprises to enforce fairness in their model's outcome by analyzing transactions in production and finding biased behavior by the model

It pinpoints the source of bias and actively mitigates the biases found in production environment

## Value:

- Automatically recommend common protected attributes to monitor during production
- Detect biases in runtime in order to catch impacts on business applications and compliance requirements without time consuming, manual data analysis
- Metrics and data to help data scientists further troubleshoot issues in data sets or models
- Mitigate biases in runtime in order to enforce regulatory or enterprise fairness guardrails in real time



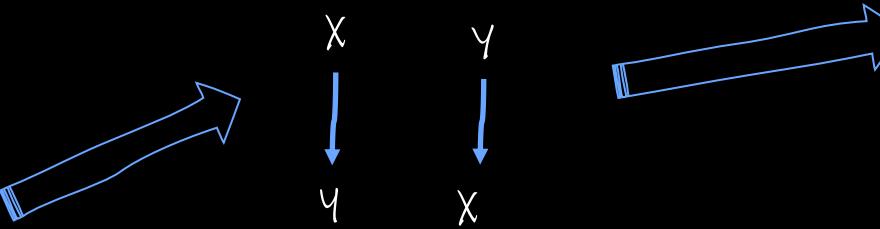


$$\text{disparate impact ratio} = (0 / X) / (Y / Y) = 0\%$$

So, according to DI ratio, Model is highly biased.

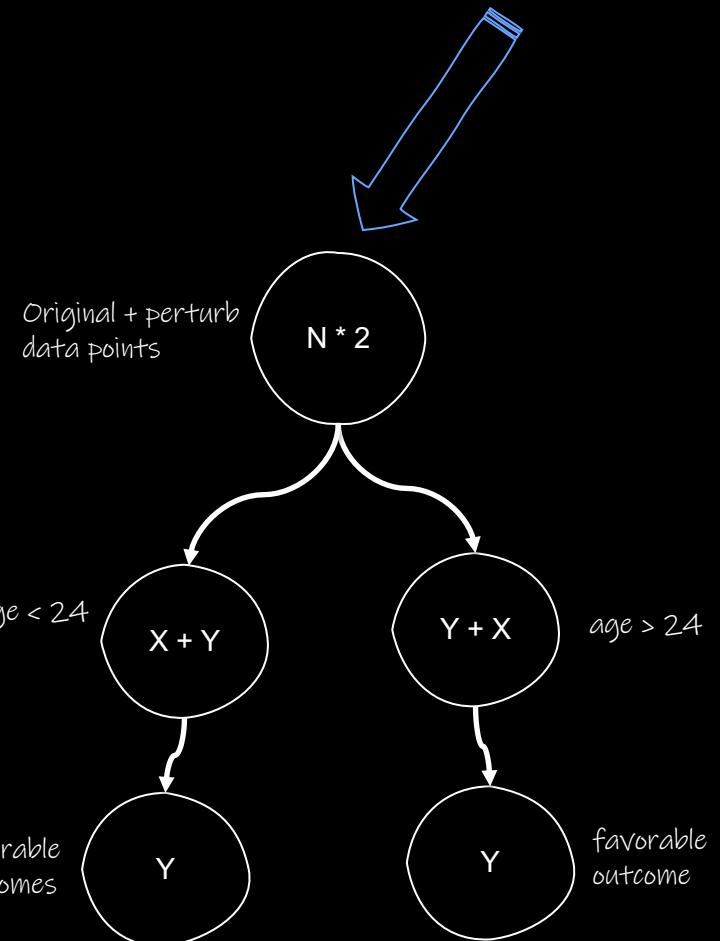
biased ??

Perturbation



Perturbed records scoring

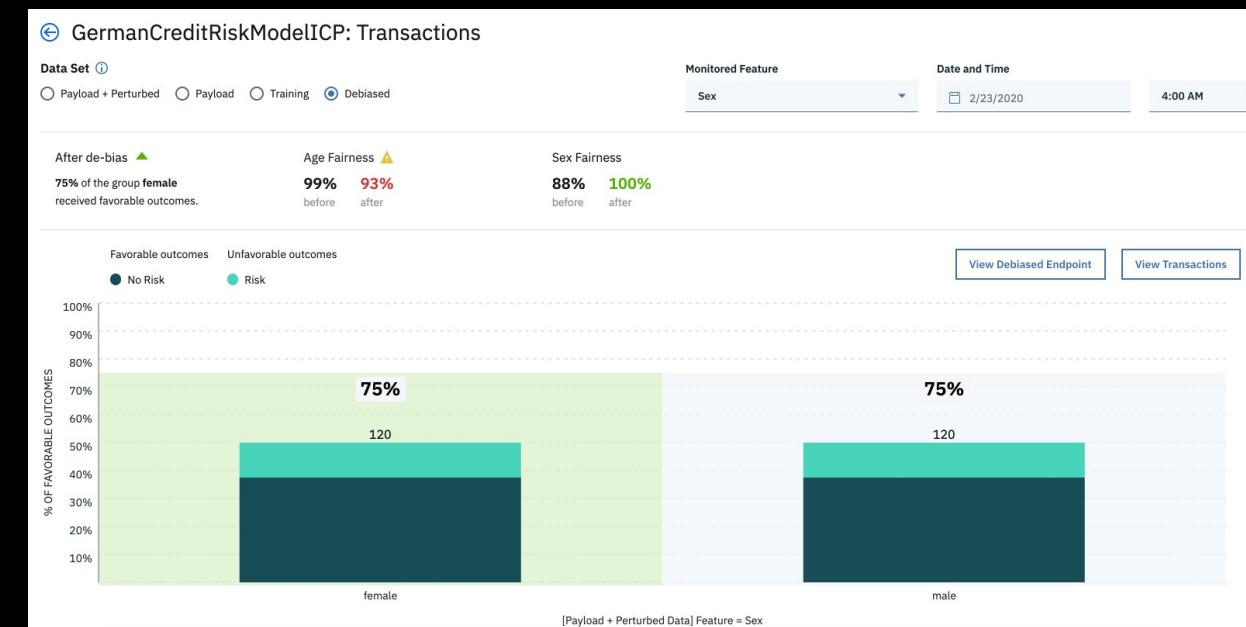
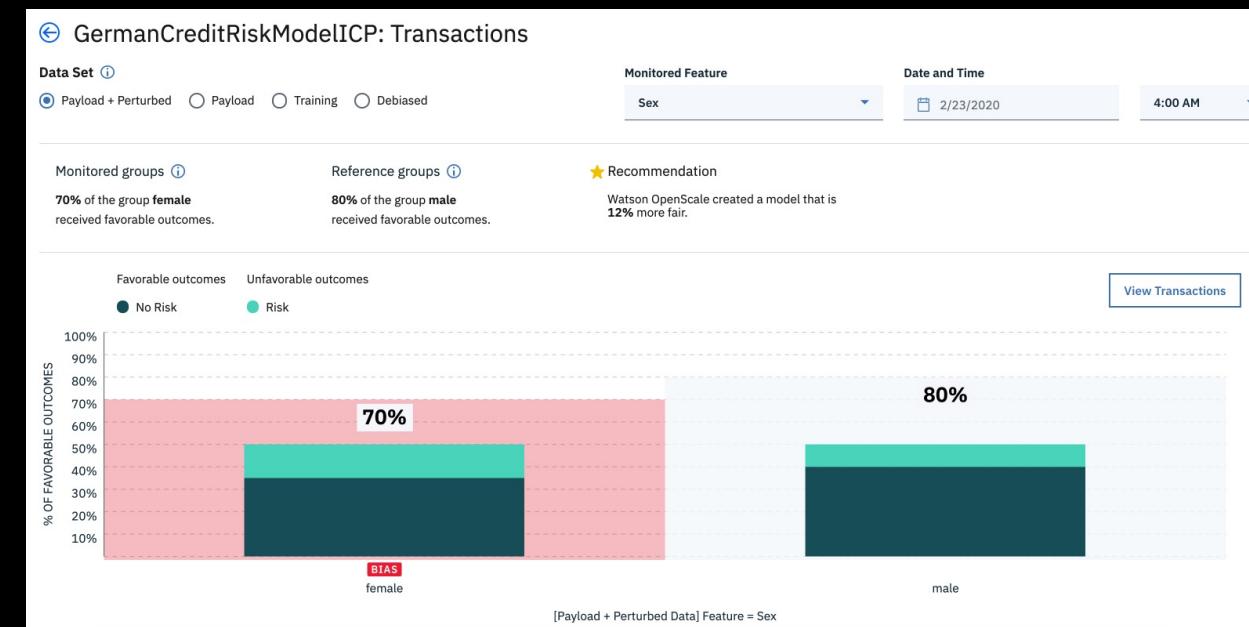
let's say  
No change in prediction



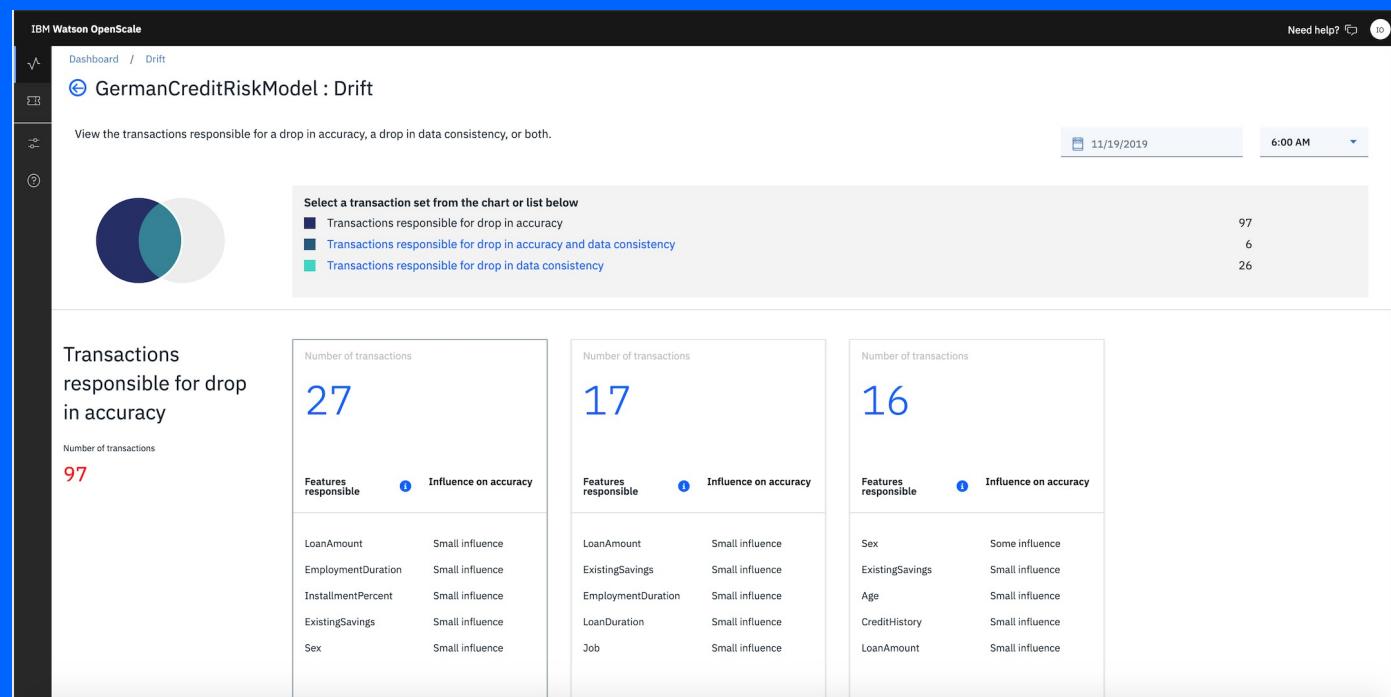
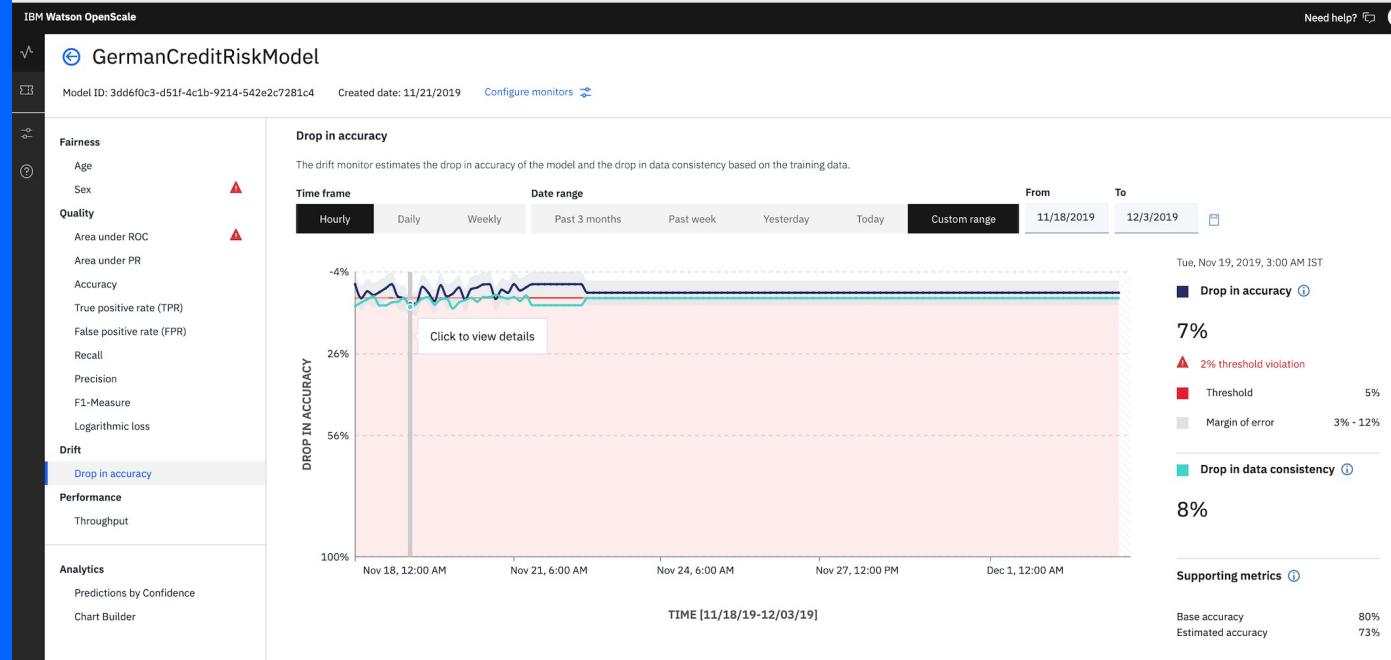
$$\text{disparate impact ratio} = (Y / (X + Y)) / (Y / (Y + X)) = 100\%$$

You see..

# Bias Mitigation – Original Output      Bias Mitigation – De-biased output



# Watson OpenScale will automatically detect drifted transactions and pinpoint datapoints that contribute to drift



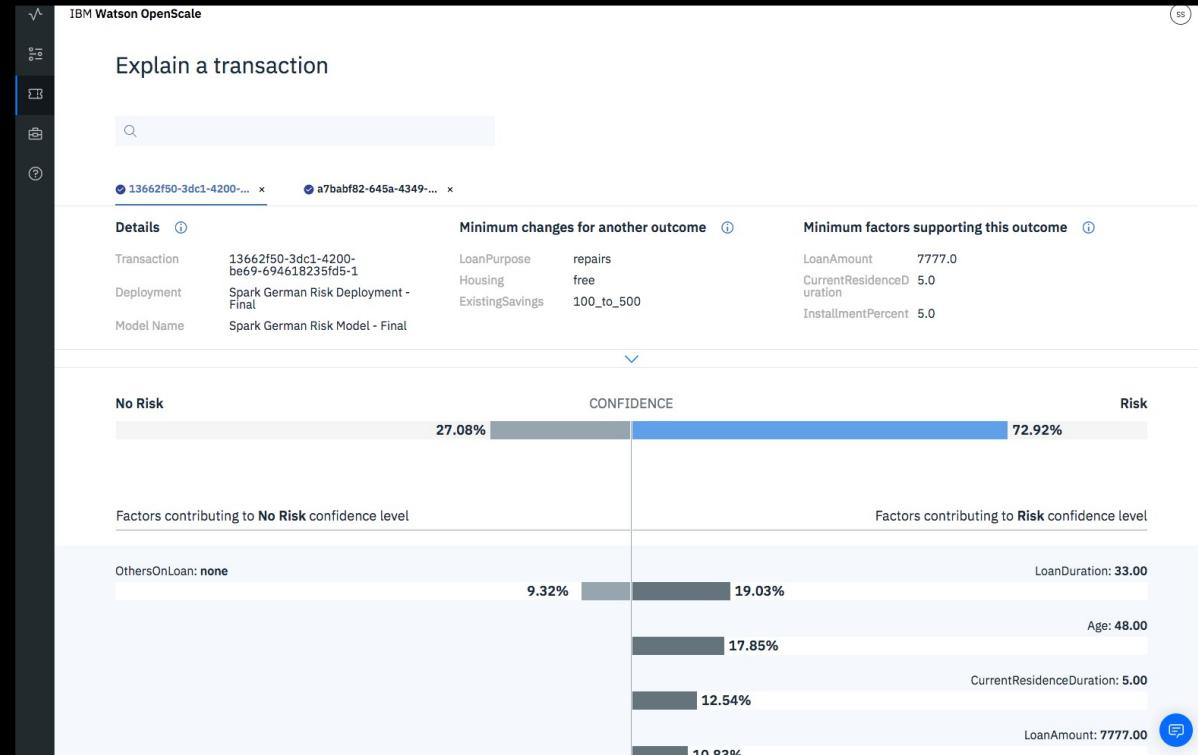
# Explainability

OpenScale records every individual transaction and drills down into its working to explain how the model makes decisions

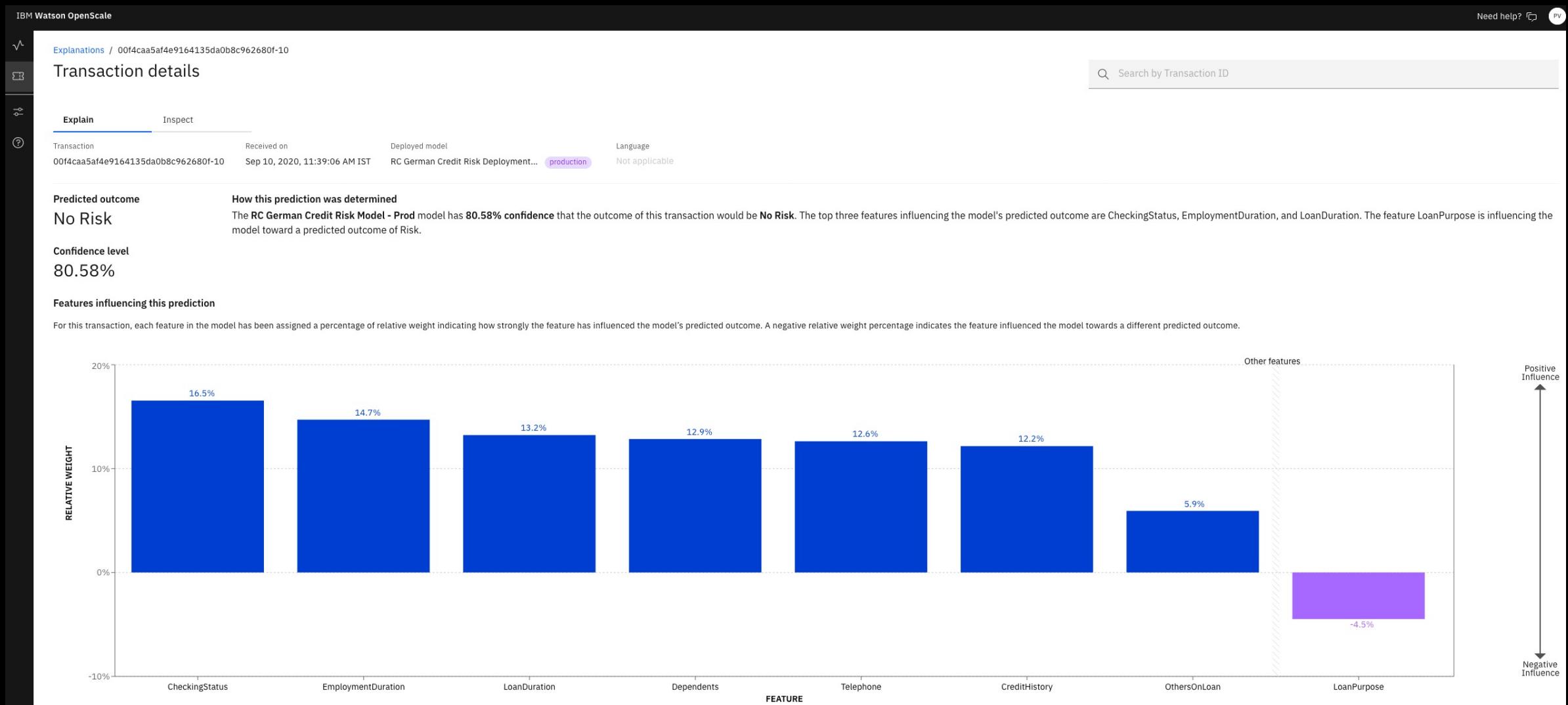
It provides a simple explanation that is user friendly and interactive

## Value:

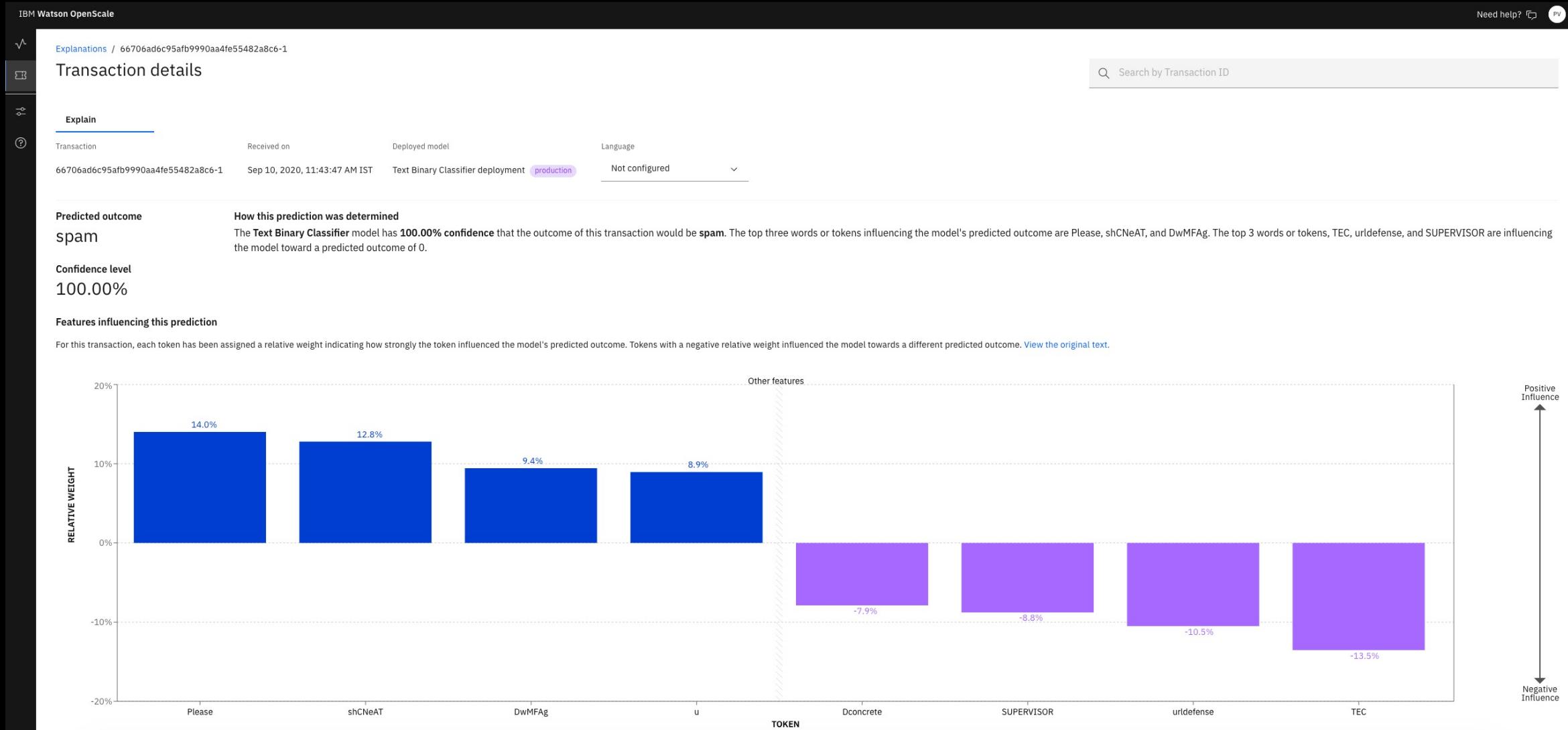
- Explain individual transaction level decisions made by the model in run time, including details about most important attributes and their values in order to assist in compliance and customer care situations
- Analyze individual transactions in a what-if manner in order to understand how model behavior will change in different business situations



# Lime Explanation for structured data



# Explanation for unstructured text



# Explanation for unstructured image

IBM Watson OpenScale Need help? PV

Explanations / 29c1e2acd1fb469ea23c5ca247a72f16-1

Transaction details

Search by Transaction ID

Explain

Transaction Received on Deployed model Language  
29c1e2acd1fb469ea23c5ca247a72f16-1 Sep 10, 2020, 12:20:52 PM IST Fashion MNIST Model Deployment production Not applicable

Predicted outcome How this prediction was determined  
2 The Fashion MNIST Model model has **96.84% confidence** that the outcome of this transaction is 2. The top image regions influencing the model's predicted outcome are displayed here.

Confidence level  
**96.84%**

Features influencing this prediction  
For this transaction, the image was separated into segments. Each segment was altered to measure influence of the section on the model outcome. Sections that influenced the model's predicted outcome are presented as positive zones. Sections that influenced the model towards a different predicted outcome are presented as negative zones.



Origin Image Positive zones Negative zones

# Contrastive Explanation

Contrastive Explanation generates two explanations:

– Pertinent Positive:

- Least interesting value where decision will not change as compared to the data points prediction.
- Least interesting value is that which is very common in the domain.
- In Pertinent Positive, we try to find a data point which is as close to the “least interesting value” i.e., the median value for all features but still has the same prediction as the data point.
- PP explains the behaviour of the model as we move towards the median value of the features

– Pertinent Negative:

- Least extreme value where decision will change.
- This explains the behaviour of the model as we move away from the median value for the features.
- We stop at the first possible position where the prediction changes.

# Contrastive Explanation

IBM Watson OpenScale Need help? PV

Explanations / 00f4caa5af4e9164135da0b8c962680f-10

Transaction details

Explain **Inspect**

Reaching a different predicted outcome

For the model to have predicted a different outcome for this transaction, the value of all listed features would need to change to the indicated minimum value. Note that changing a feature value by more than the minimum value may affect the minimum change of other features for the model to predict a different outcome. Higher feature importance numbers indicate a greater likelihood of changing the prediction.

Analyze controllable features only

Feature	Original value	New value	Value for a different outcome	Importance	
Sex	male	male	female	1.00	
CheckingStatus	no_checking	no_checking	no_checking	0.00	
LoanDuration	25	25	25	0.00	
CreditHistory	prior_payments_delayed	prior_payments_delayed	prior_payments_delayed	0.00	
LoanPurpose	furniture	furniture	furniture	0.00	
LoanAmount	7279	7279	7279	0.00	
ExistingSavings	100_to_500	100_to_500	100_to_500	0.00	
EmploymentDuration	4_to_7	4_to_7	4_to_7	0.00	
InstallmentPercent	4	4	4	0.00	
Predicted outcome <b>No Risk</b>	Confidence <b>80.58%</b>	Predicted outcome <b>No Risk</b>	Confidence <b>80.58%</b>	Predicted outcome <b>Risk</b>	Confidence <b>73.81%</b>

# An Example

## Data Point:

- Input Features: Number of married years = 5, Salary = \$155K, Region=NY, Age
- Prediction: Loan = Partially Granted

**LIME Explanation:** Important features: Number of married years, Salary, Age

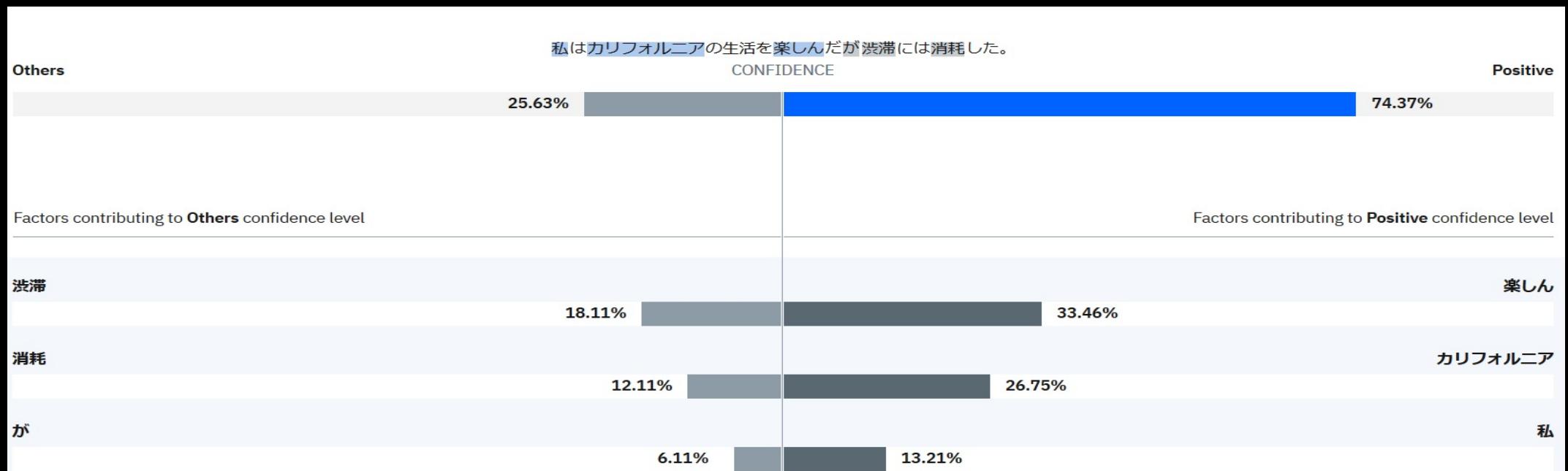
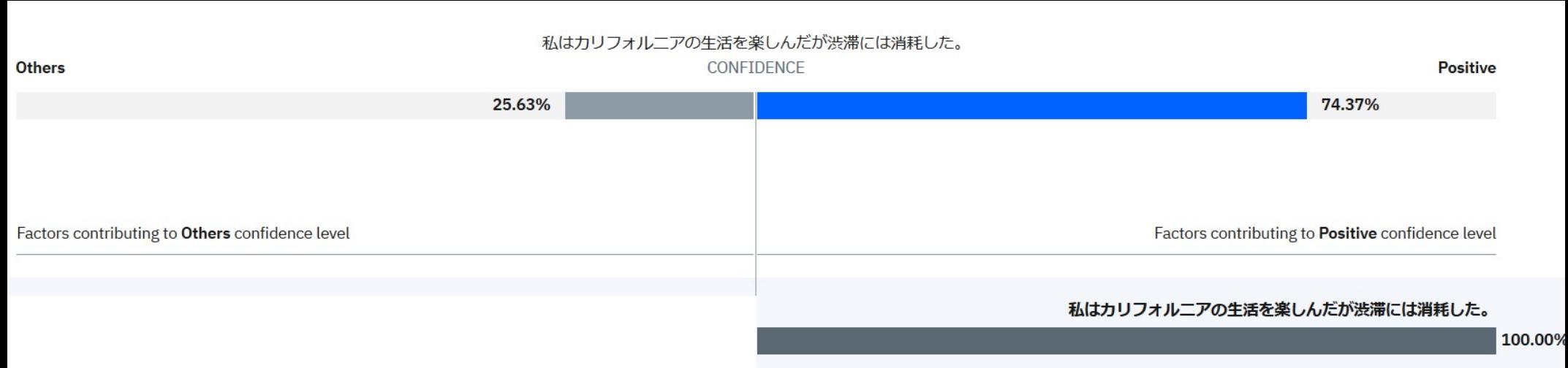
## Contrastive Explanation:

- Pertinent Negative:
  - If Number of married years was 7 and salary was in the range \$190-210K, then Loan=Granted
- Pertinent Positive:
  - Even if Number of married years = 3 and salary was in the range \$110-130K, outcome would have been Loan=Partially Granted

# Improvements to open-source LIME

- Local explanations for certain datapoints tend to be difficult to interpret with the default lime. OpenScale LIME integration detects such scenarios and improves the explanation by using a custom regressor with custom constraints for local linear approximation of the model behavior.
- OpenScale has made an improvement to sample the perturbations needed for generating explanations once per model and then cache those. This results in scoring cost for the first explanation and subsequently zero cost for further explanations.
- OpenScale has made improvements to LIME to tolerate scoring failures till a configurable threshold percentage of all scoring requests.
- Support for Chinese, Japanese, Korean languages in unstructured text explanations.

# Explanation for Japanese text



# Emerging use case: Model Risk Management in Financial Services

## Problem:

Current risk management practices are not optimized for AI

- **Open-source frameworks** not supported
- Additional **data science skills** required for validating AI models
- Processes rely on **manual interaction** between validator and model developer
- Current systems **focus more on the documentation and workflow** aspects of model validation, **no active testing**
- No focus on **active production monitoring**

Watson OpenScale will **automate active testing** of models for **validation and monitoring** and **synchronize results with governance platforms**

Dashboard /

## ⬅ Credit Risk pre-prod deployment in B2 Evaluations

[Actions](#) ▾

## Model

Credit Risk pre-prod deployment in B2 Pre-production

## Description

Credit Risk pre-prod deployment in B2

## Model ID

21b4fe7b-5801-4266-9190-609ffe449646



## Tests run

**3**● Tests passed**3**● Tests failed**0**

## Evaluation date

Wed, Nov 27, 2019, 4:17 PM IST

## Test data set

german\_credit\_data\_biased\_test.csv

## Number of test records

200

## Number of explanations

10

## Fairness

**95.10%**Green within threshold

100 records evaluated

[Configure](#)

## Quality

**0.74**Green within threshold

100 records evaluated

[Configure](#) [Configure](#)

## Drift

**1.30%**Green within threshold

100 records evaluated

[Configure](#)

## Fairness by feature

Sex 95.10%

Age 95.20%

## Quality metrics

True positive rate (TPR) 0.58

Area under ROC 0.74

Precision 0.74

F1-Measure 0.65

Accuracy 0.81

Logarithmic loss 0.42

False positive rate (FPR) 0.09

Area under PR 0.65

Recall 0.58

## Drift metrics

Drop in data consistency 7.75%

Drop in accuracy 1.30%

Predicted accuracy 78.20%

Base accuracy 79.50%

# IBM Watson OpenScale + IBM OpenPages

Model Risk Management Solution for Financial Services

## IBM OpenPages

### Model Risk Governance

Store, manage and monitor a comprehensive model inventory



## IBM Watson OpenScale

Active testing for model validation and continuous monitoring

# Thank you

