QuantUniversity, LLC

www.quantuniversity.com

MathWorks

Scalable Data Science Pipelines with QuSandbox & MATLAB Online Server

Presented By:

Sri Krishnamurthy, CFA, CAP

sri@quantuniversity.com

www.quantuniversity.com

MathWorks Computational Finance Conference 2021

Online

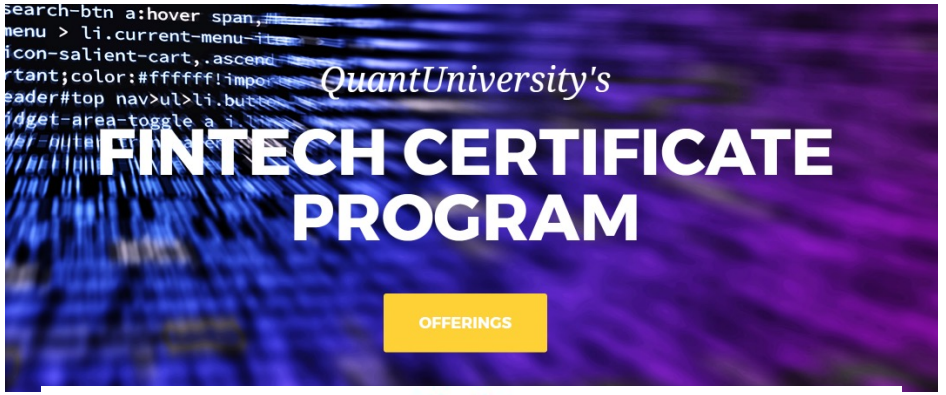
Speaker bio

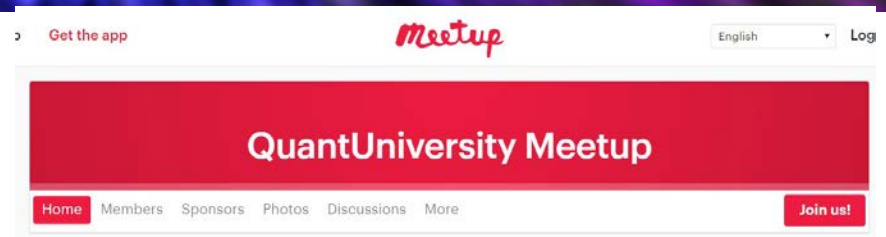
Sri Krishnamurthy
Founder and CEO
QuantUniversity

- Advisory and Consultancy for Financial Analytics
- Prior Experience at MathWorks, Citigroup and Endeca and 25+ financial services and energy customers.
- Columnist for the Wilmott Magazine
- Author of forthcoming book "Pragmatic AI and ML in Finance"
- Teaches AI/ML and Fintech Related topics in the MS and MBA programs at Northeastern University, Boston
- **Reviewer:** Journal of Asset Management

QuantUniversity

- Boston-based Data Science, Quant Finance and Machine Learning training and consulting advisory
- Trained more than 1000 students in Quantitative methods, Data Science and Big Data Technologies using MATLAB, Python and R
- Building QuSandbox a platform for AI and Machine Learning Experimentation





QuantUniversity, LLC
www.quantuniversity.com

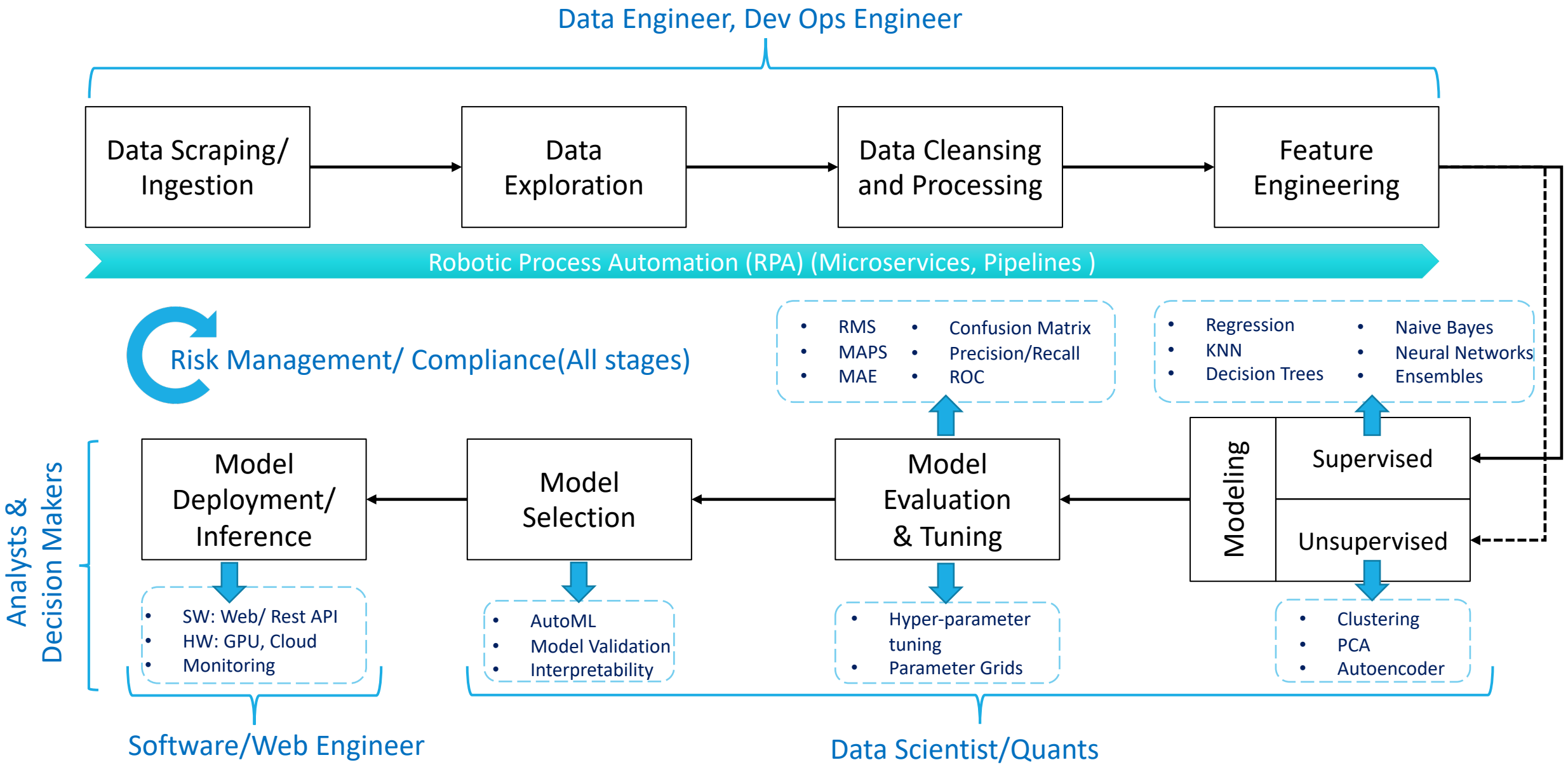
Agenda

1. Data Science Pipelines & QuSandbox
2. Case study:
   - NLP Pipeline for Sentiment Analysis of EDGAR filings

QuantUniversity, LLC
www.quantuniversity.com

Pipelines and the QuSandbox

Machine Learning Workflow



QuantUniversity, LLC
www.quantuniversity.com

The QU approach

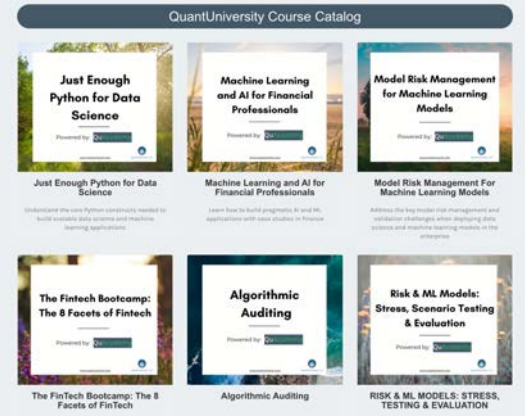
QuSandbox

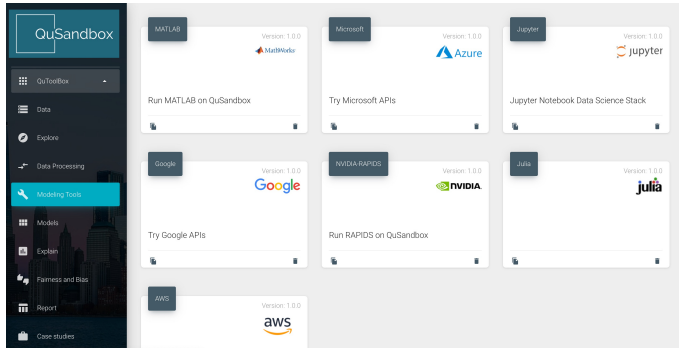
Request DEMO at
info@qusandbox.com

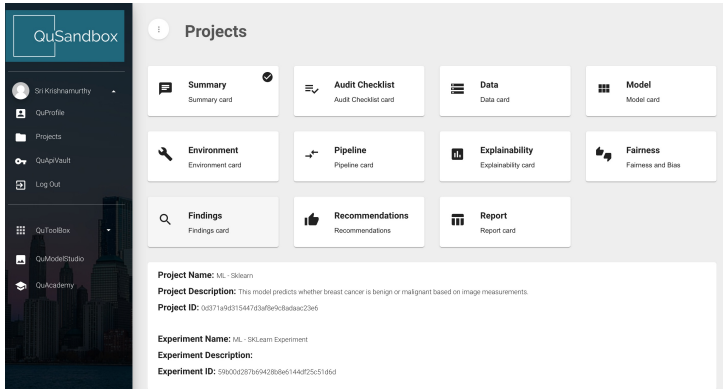
EDUCATION
QUACADEMY

EXPERIMENTATION
QUTOOLBOX

ENABLEMENT
QUSANDBOX







Filter By Courses

Risk and ML Models

Filter By Topics

All

Search Project

1. Introduction to ML, AI and Risk

Duration: 1 hr *QUcredits: 1*

Introduction to Machine Learning, AI and Risk

MORE DETAILS

2. Stress testing and Scenario generation

Duration: 1 hr *QUcredits: 1*

Introduction to Machine Learning, AI and Risk

MORE DETAILS

3. Metrics and Evaluation for risk in models

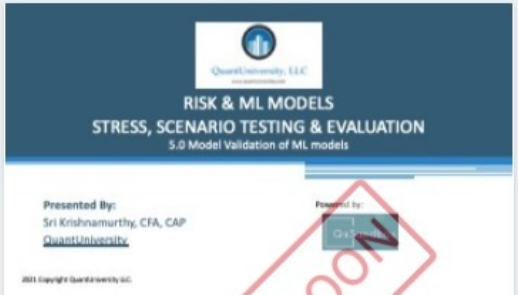
Duration: 1 hr *QUcredits: 1*

Introduction to Machine Learning, AI and Risk

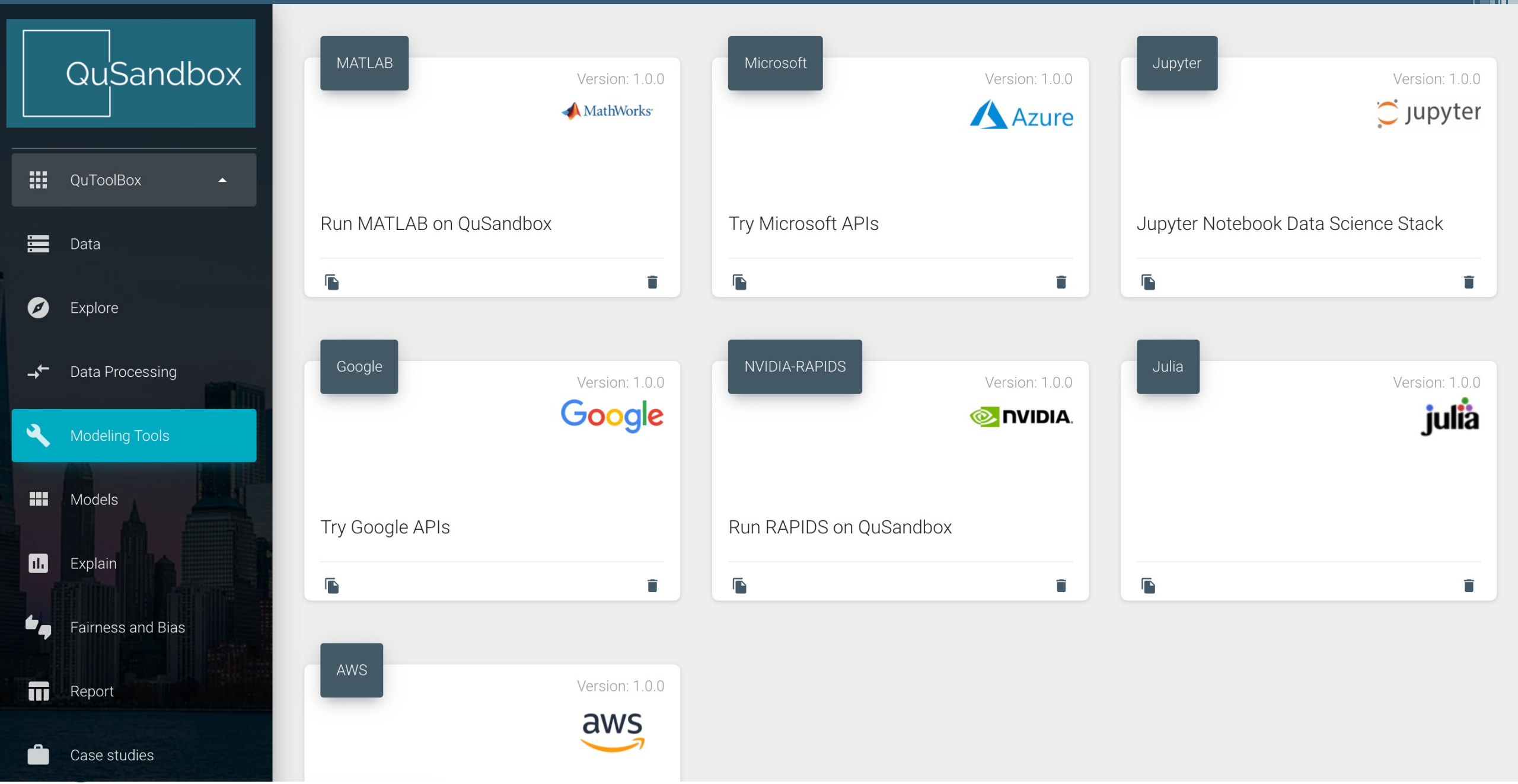
MORE DETAILS

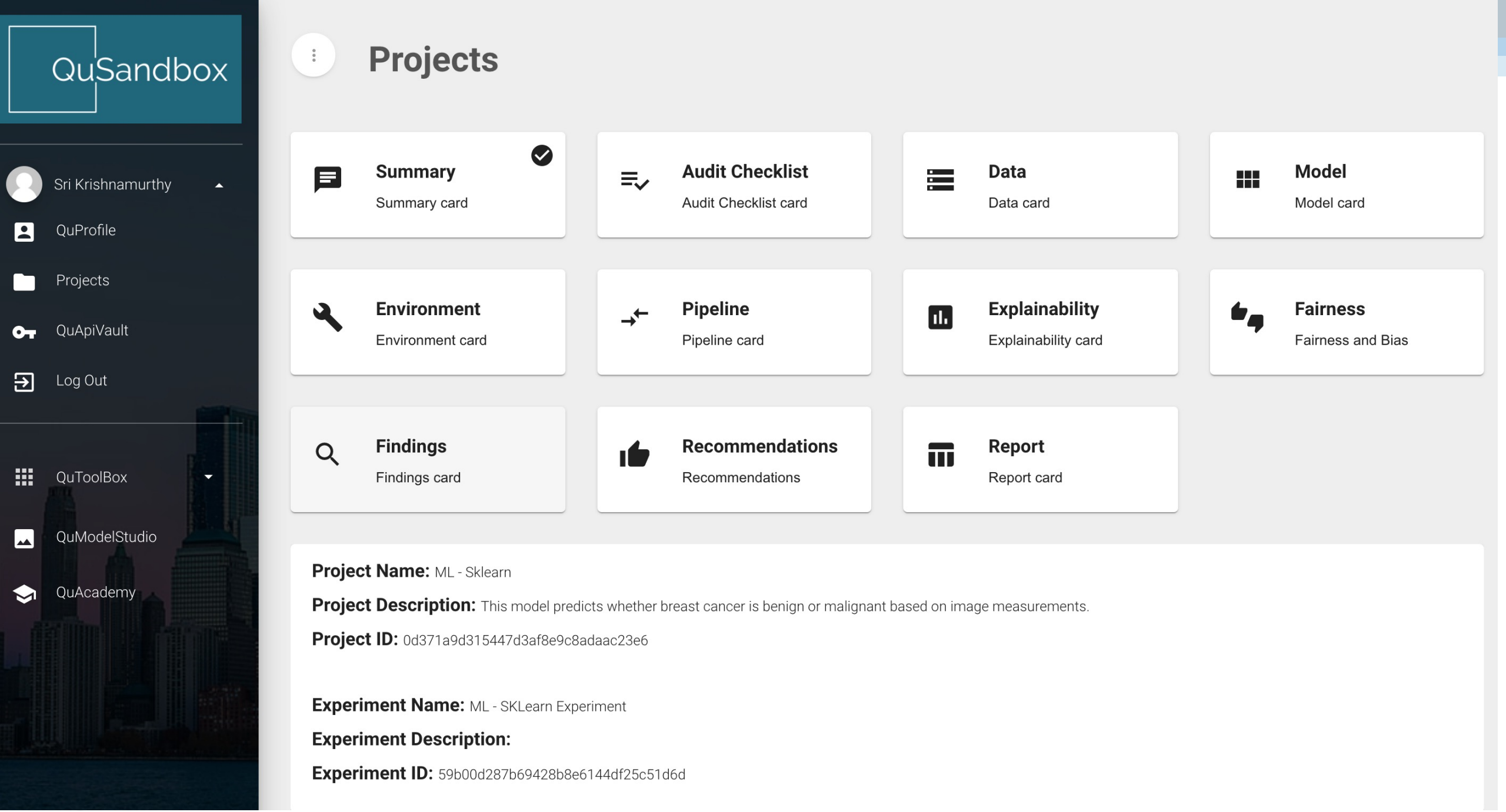
4. Anomalies and Outliers

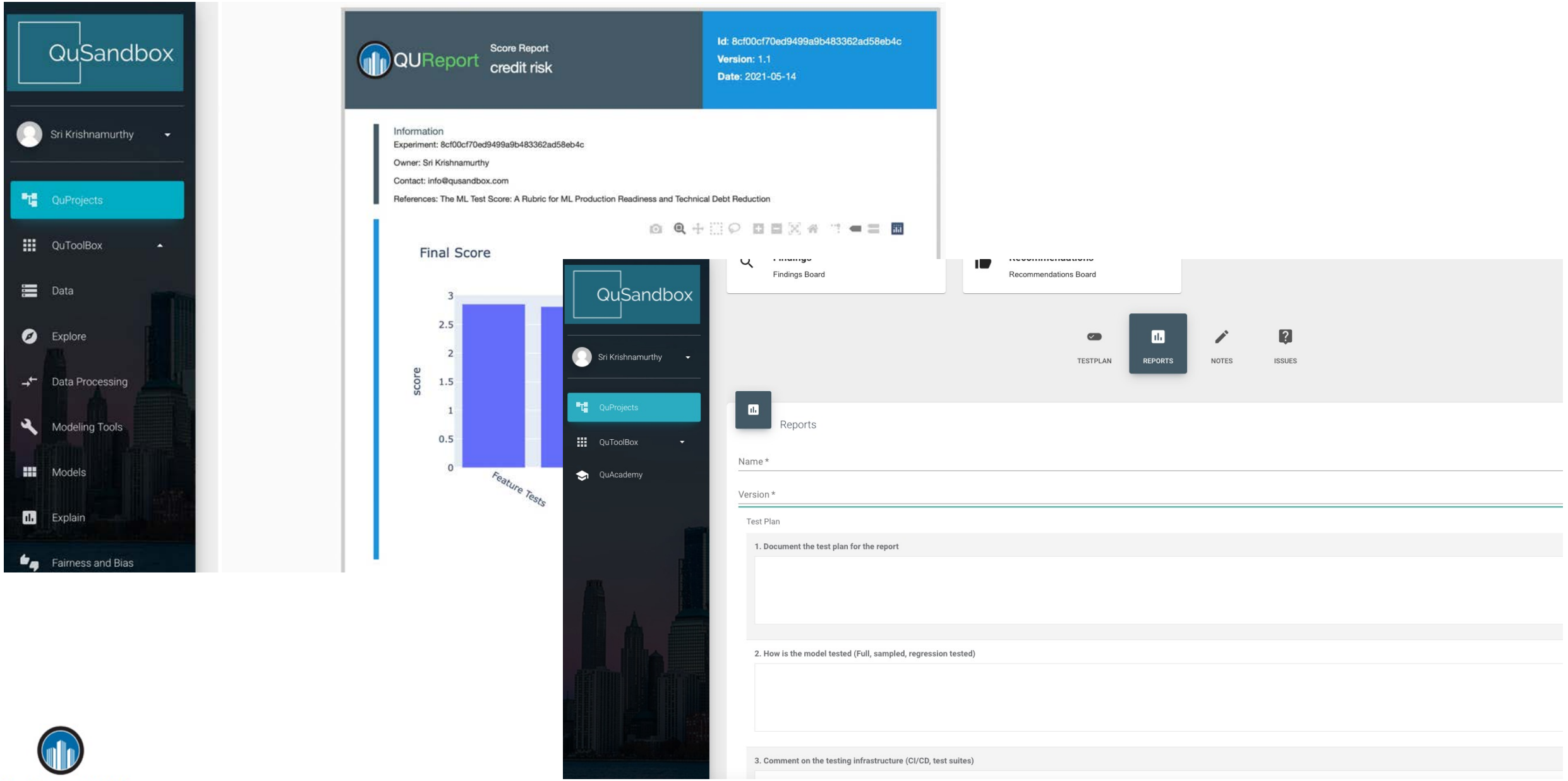
5. Model Validation of ML



6. Frontier topics and Wrap


QuSandbox
QuToolBox
Data
Explore
Data Processing
Modeling Tools
Models
Explain
Fairness and Bias
Report
Case studies
MATLAB
Version: 1.0.0
MathWorks
Run MATLAB on QuSandbox
Microsoft
Version: 1.0.0
Azure
Try Microsoft APIs
Jupyter
Version: 1.0.0
jupyter
Jupyter Notebook Data Science Stack
Google
Version: 1.0.0
Google
Try Google APIs
NVIDIA-RAPIDS
Version: 1.0.0
NVIDIA
Run RAPIDS on QuSandbox
Julia
Version: 1.0.0
julia
AWS
Version: 1.0.0
aws


QuSandbox
Sri Krishnamurthy
QuProfile
Projects
QuApiVault
Log Out
QuToolBox
QuModelStudio
QuAcademy
Projects
Summary
Summary card
Audit Checklist
Audit Checklist card
Data
Data card
Model
Model card
Environment
Environment card
Pipeline
Pipeline card
Explainability
Explainability card
Fairness
Fairness and Bias
Findings
Findings card
Recommendations
Recommendations
Report
Report card
Project Name: ML - Sklearn
Project Description: This model predicts whether breast cancer is benign or malignant based on image measurements.
Project ID: 0d371a9d315447d3af8e9c8adaac23e6
Experiment Name: ML - SKLearn Experiment
Experiment Description:
Experiment ID: 59b00d287b69428b8e6144df25c51d6d


QuSandbox
Sri Krishnamurthy
QuProjects
QuToolBox
Data
Explore
Data Processing
Modeling Tools
Models
Explain
Fairness and Bias
QUReport
Score Report
credit risk
Id: 8cf00cf70ed9499a9b483362ad58eb4c
Version: 1.1
Date: 2021-05-14
Information
Experiment: 8cf00cf70ed9499a9b483362ad58eb4c
Owner: Sri Krishnamurthy
Contact: info@qusandbox.com
References: The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction
Final Score
score
Feature Tests
Findings Board
Recommendations Board
TESTPLAN
REPORTS
NOTES
ISSUES
QuSandbox
Sri Krishnamurthy
QuProjects
QuToolBox
QuAcademy
Reports
Name *
Version *
Test Plan
1. Document the test plan for the report
2. How is the model tested (Full, sampled, regression tested)
3. Comment on the testing infrastructure (CI/CD, test suites)

Case study

Goal

- Understanding sentiments in Earnings call transcripts

CORPORATE PARTICIPANTS

Dana Quattrochi athenahealth, Inc. - IR
Jonathan Bush athenahealth, Inc - Chairman and CEO
Tim Adams athenahealth, Inc - CFO
Andy Hurd Epocrates - President and CEO
Rob Cosinuke athenahealth, Inc. - Chief Marketing Officer

CONFERENCE CALL PARTICIPANTS

Sean Wieland Piper Jaffray & Co. - Analyst
Jamie Stockton Wells Fargo Securities, LLC - Analyst
George Hill Citigroup - Analyst
Greg Bolan Sterne, Agee & Leach - Analyst
Ryan Daniels William Blair & Company - Analyst
Rich Close Avondale Partners - Analyst
Sandy Draper Raymond James - Analyst
David Bayer Northland Securities - Analyst
Dave Windley Jefferies & Co. - Analyst
Charles Rhyee Cowen and Company - Analyst
Bret Jones Oppenheimer & Co. - Analyst
Michael Cherny ISI Group - Analyst
Tony Bartsch Park West Asset Management - Analyst

PRESENTATION

Operator

Welcome to the athenahealth conference call. I would now like to turn the call over to Ms. Dana Quattrochi. You may now begin.

Dana Quattrochi - *athenahealth, Inc. - IR*

Good morning and thank you for joining us. With me on the call today is Jonathan Bush, our Chairman and CEO; Tim Adams, our Chief Financial Officer; Rob Cosinuke, our Chief Marketing Officer; and Andy Hurd, President and CEO of Epocrates.

Challenges

- Interpreting emotions
- Labeling data

NLP pipeline

Stage 1

Data Ingestion from Edgar

Stage 2

Pre-Processing

Stage 3

Invoking APIs to label data

- Amazon Comprehend API
- Google API
- Watson API
- Azure API

Stage 4

Compare APIs

Stage 5

Build a new model for sentiment Analysis

QUSANDBOX

THE NEW DECALOGUE

ML GOVERNANCE IN THE AGE OF DATA SCIENCE AND AI
WWW.QUANTUNIVERSITY.COM

OUR SERVICES

- Model Governance and Algorithmic Audits
- Model Life Cycle Management
- AI/ML Onboarding
- Third-party Model Validation
- Training and Education
- QuSandbox

ABOUT US

QuantUniversity is a quantitative analytics and model risk advisory based in Boston, MA. We provide Quant Finance, Data Science and Machine Learning based solutions focused on model-risk made accessible through QuSandbox. QuantUniversity has worked with analysts and executives from Bloomberg, Fidelity, Ford, Goldman Sachs, IBM, J.P. Morgan Chase, Nataxis Global Advisors, Pan Agora, T.D. Securities, and other institutions, providing quantitative advisory services, analytics training, and model risk solutions. Contact us at www.quantuniversity.com

10 THINGS YOU NEED TO KNOW ABOUT MODEL GOVERNANCE FOR AI/ML MODELS

1. DEFINING MODELS
Models are not just restricted to code and associated parameters. You have to factor data, the programming environment and packages, parameters and hyperparameters along with the model code.

2. GOVERNING MACHINE LEARNING MODELS
You could have hundreds of machine learning models working alongside traditional models. A comprehensive framework is needed to factor the nuances of machine learning models in your governance process.

3. MODEL VERIFICATION AND VALIDATION OF MACHINE LEARNING MODELS
It's not just sufficient to verify if machine learning models work with historical test/validation datasets from a technical perspective. You have to validate if the models can be used for business decision making.

4. PERFORMANCE METRICS AND EVALUATION CRITERIA
The choice of performance metrics and evaluation criteria depends on how the models would be used and for what purpose. Evaluate the choices carefully.

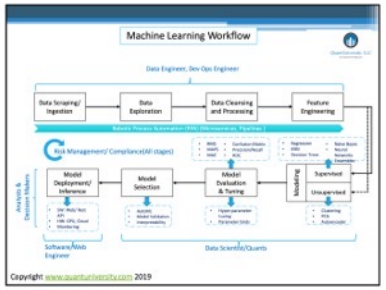
5. MODEL INVENTORY AND TRACKING
Avoid model "clutter" by having a formal model inventory and tracking system. You need to track models, data snapshots, parameters, hyperparameters, programming environments etc. In addition, the entire pipeline needs to be tracked. Provenance tracking is important for reproducibility.

INFO@QUSANDBOX.COM
WWW.QUSANDBOX.COM
617-283-7904

QU

THE NEW DECALOGUE

ML GOVERNANCE IN THE AGE OF DATA SCIENCE AND AI
WWW.QUANTUNIVERSITY.COM



QuantUniversity, LLC
www.quantuniversity.com

10 THINGS YOU NEED TO KNOW ABOUT MODEL GOVERNANCE FOR AI/ML MODELS (CONT'D)

6. DATA GOVERNANCE AND MODEL GOVERANANCE
Machine Learning models are by design data driven. Integrating Data governance and model governance aspects is essential.

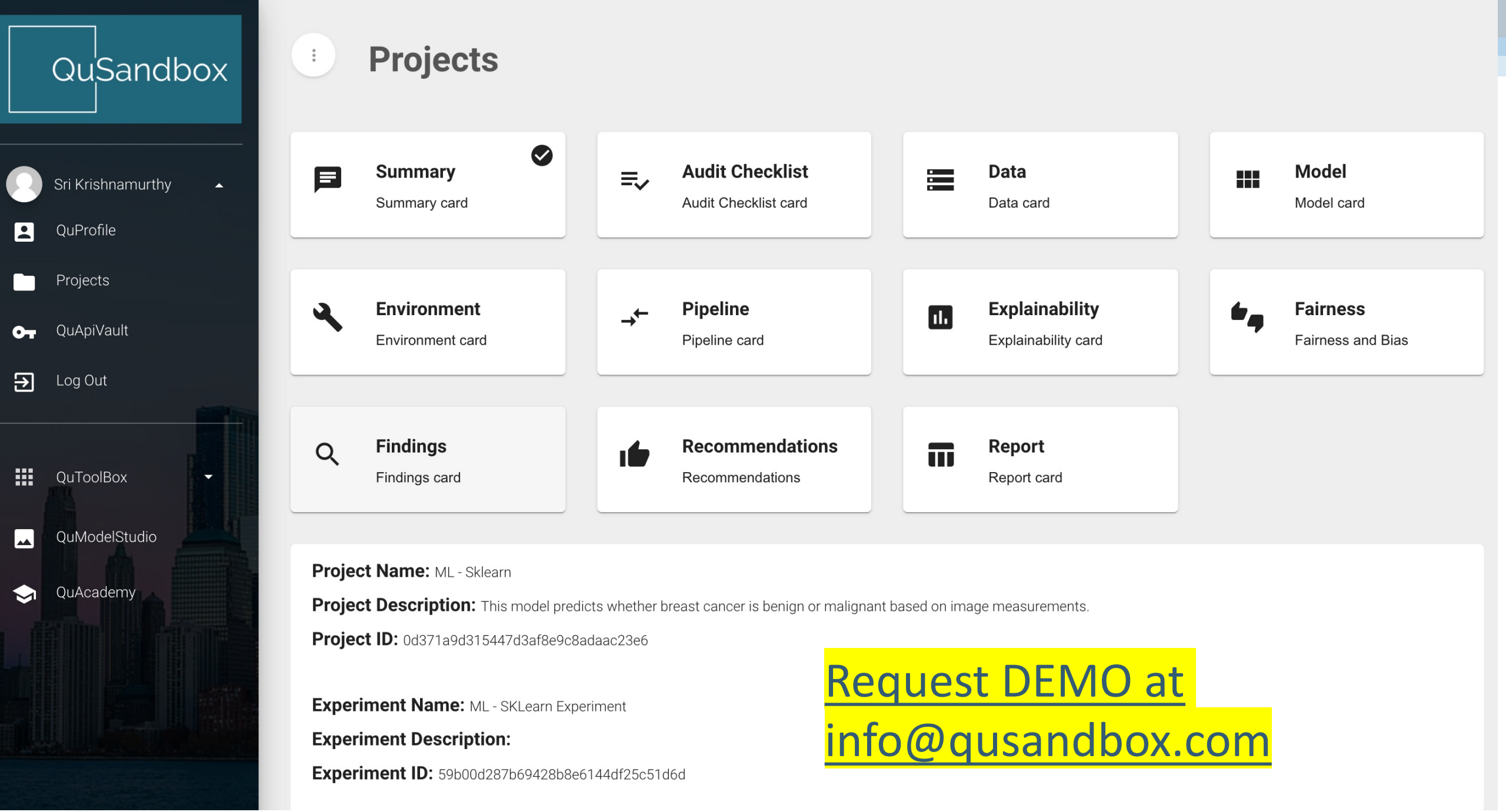
7. DEVELOPMENT MODELS VS PRODUCTION MODELS
As you design models for inference, scalability, performance considerations need to be factored. Models may have to be redesigned/compiled to factor production requirements. It is important to test models to ensure production models behave as they were designed.

8. FAIRNESS, REPRODUCIBILITY, AUDITABILITY, EXPLAINABILITY, INTERPRETABILITY, BIAS
Depending on the application, models should be evaluated to ensure Fairness, Reproducibility, Auditability, Explainability, Interpretability & Bias considerations are met.

9. MACHINE LEARNING CHOICES
As the field of machine learning matures, you have multiple options, Automatic Machine Learning, ML as a service, Pre-trained models and models developed from scratch etc. bringing different model governance considerations.

10. ROLES AND RESPONSIBILITIES
With AI and ML making strides, you have many new roles in your model building workflow. (Data engineers, scientists, model evaluators, cloud engineers, DevOps, MLOps etc.). Factor the new roles and define clear responsibilities for all the key stakeholders in the model lifecycle.

INFO@QUSANDBOX.COM
WWW.QUSANDBOX.COM
617-283-7904


QuSandbox
Sri Krishnamurthy
QuProfile
Projects
QuApiVault
Log Out
QuToolBox
QuModelStudio
QuAcademy
Projects
Summary
Summary card
Audit Checklist
Audit Checklist card
Data
Data card
Model
Model card
Environment
Environment card
Pipeline
Pipeline card
Explainability
Explainability card
Fairness
Fairness and Bias
Findings
Findings card
Recommendations
Recommendations
Report
Report card
Project Name: ML - Sklearn
Project Description: This model predicts whether breast cancer is benign or malignant based on image measurements.
Project ID: 0d371a9d315447d3af8e9c8adaac23e6
Experiment Name: ML - SKLearn Experiment
Experiment Description:
Experiment ID: 59b00d287b69428b8e6144df25c51d6d
Request DEMO at
info@qusandbox.com

QuantUniversity, LLC

www.quantuniversity.com

Thank you!

Contact

Sri Krishnamurthy, CFA, CAP
Founder and CEO
QuantUniversity LLC.

LinkedIn srikrishnamurthy

www.QuantUniversity.com

QuantUniversity, LLC

www.quantuniversity.com