



# OS-Climate NLP demo using ODH services

---

Shrey Anand  
Data Scientist

---

Karan Chauhan  
Data Scientist

# Contents

- Context
- Environment Setup
- NLP notebooks
- Results

# OS Climate

- Data and software to drive global investments toward green businesses
- Add real world impact to risk and reward parameters
- Decision making relies on data from different sources
- Aggregate, Standardize and Predict
- [Keynote by Michael Tiemann, Marcel Hild, Erik Erlandson](#)

# Context

- Dataset: Climate reports in PDF format
  - Commitments, carbon emissions, etc.
- Challenges
  - No standards followed in the reports
  - Finding information from them is cumbersome

# Context

- Solution:
  - NLP model by the Allianz group in the OS climate community
- Pipelines:
  - OSC cluster: Openshift cluster created using Operate First principles
  - Open Data Hub services like Jupyterhub and Superset deployed

# Environment Setup

- Repository: [aicoe-osc-demo](#)
- Data Science project template
- AICoE CI pipeline

# Notebook walkthrough

- Architecture diagram
- PDF text extraction
- Infer Relevance
- Infer KPI
- Results table

# Pipeline

- Elyra UI
- [Kubeflow page](#)
- Results on s3 storage



# Visualization

- [Superset dashboard](#)

# Summary

We showed how to use ODH tools to create reproducible cloud native DS environment to develop notebooks and then create a pipeline that takes raw pdfs from s3, parses them, runs the NLP model, save the results on Trino and then visualize them using Superset.

# Additional slides



