

A person is seen from the side, wearing a headset and playing a game on a dual-monitor setup. The left monitor displays a colorful, action-packed game scene, while the right monitor shows a social media or chat interface. The person is sitting in a black leather chair, and the desk is illuminated with blue and red lights. The background is dark, suggesting a gaming room or a late-night session.

Real-Time Toxicity Detection and Management in Video Gaming: A Comprehensive Data-Driven Approach

Team 13: Mike Jiang, Shivani Vats, Thomas
Farrell, Priyanka Prashanth Kumar, Yen Chen Hsu



Motivation

Decreased
Communication
Efficiency

Reduced Team
Morale and
Satisfaction

Hindered
Creativity and
Innovation

- A Worse Player Experience
- Higher Player Churn Rates

Decreased Communication Efficiency

Toxic communication severely impacts team efficiency and collaboration. Negative interactions can slow down decision-making processes and lead to misunderstandings or overlooking of critical information.



Reduced Team Morale and Satisfaction

A toxic environment negatively affects team morale. This can harm performance, alienate players, and lead to lower enjoyment and retention



Hindered Creativity and Innovation

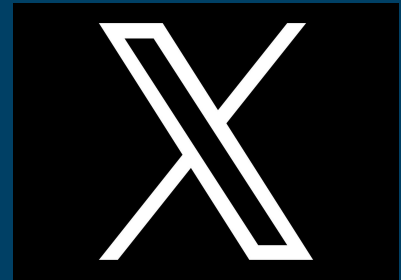
In a negative atmosphere, team members may hesitate to share ideas or strategies for fear of criticism or negative feedback, limiting the team's creative and innovative potential.



USE CASE #1 Video Games



USE CASE #2 Social Media



USE CASE #3 Internal Communications

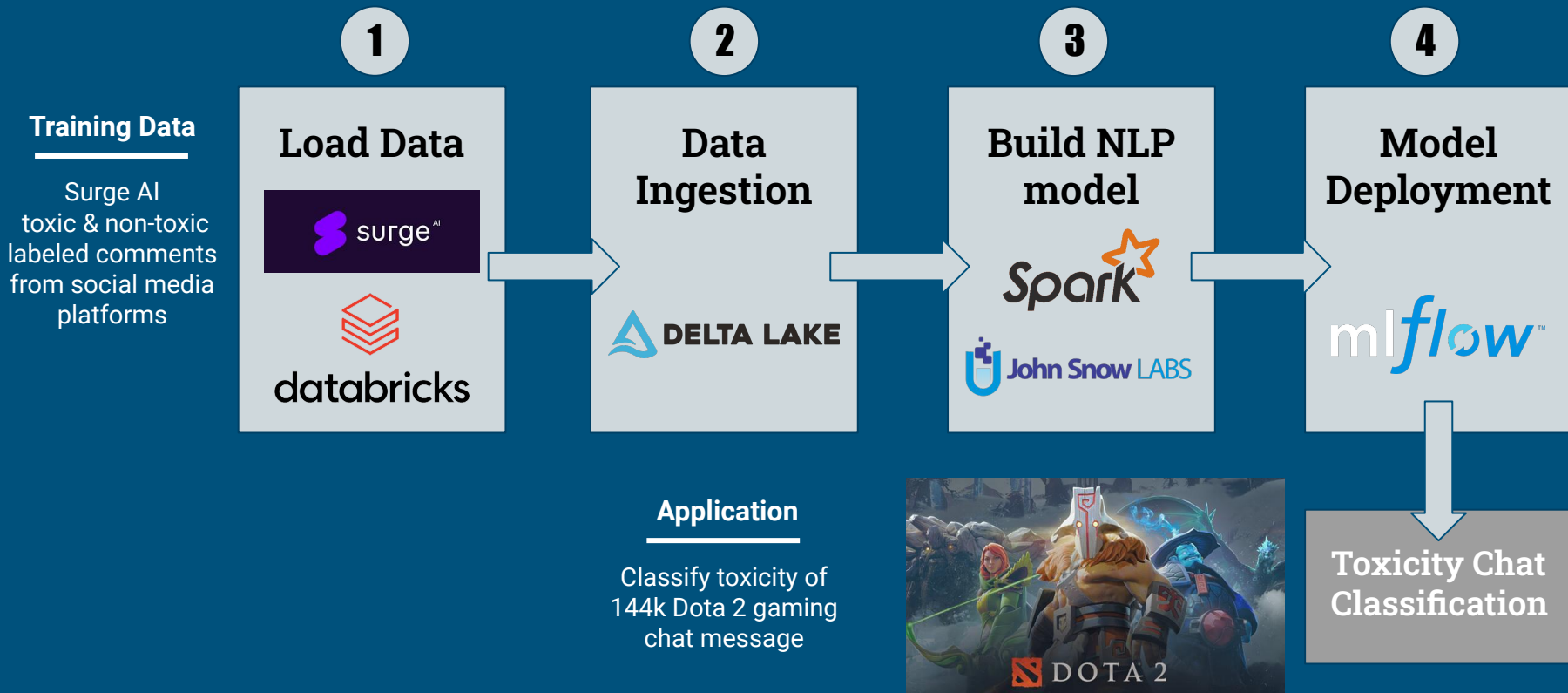


Why Databricks?

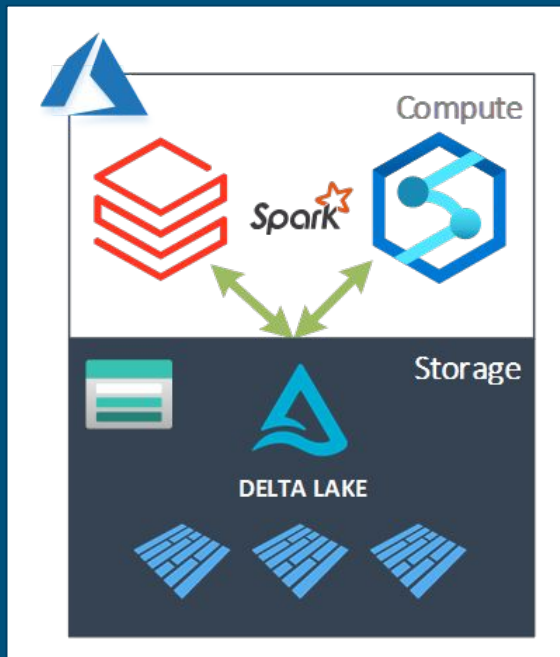
Benefits of Databricks Migration



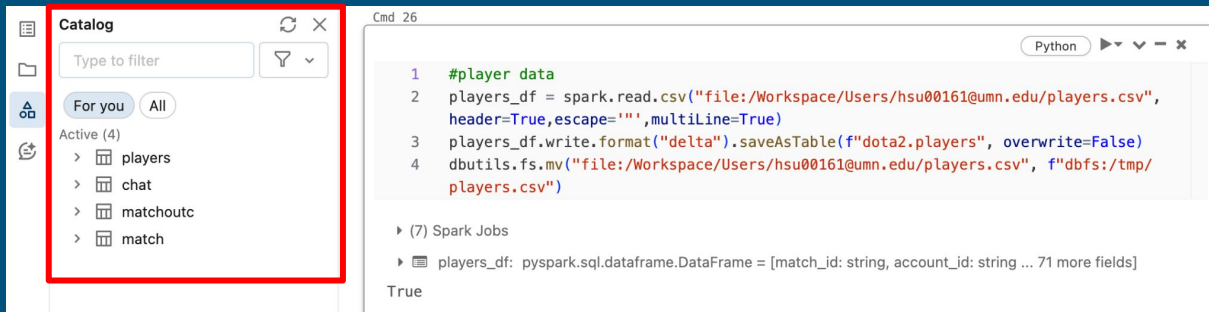
Demo: Toxicity Chat Detection on Databricks



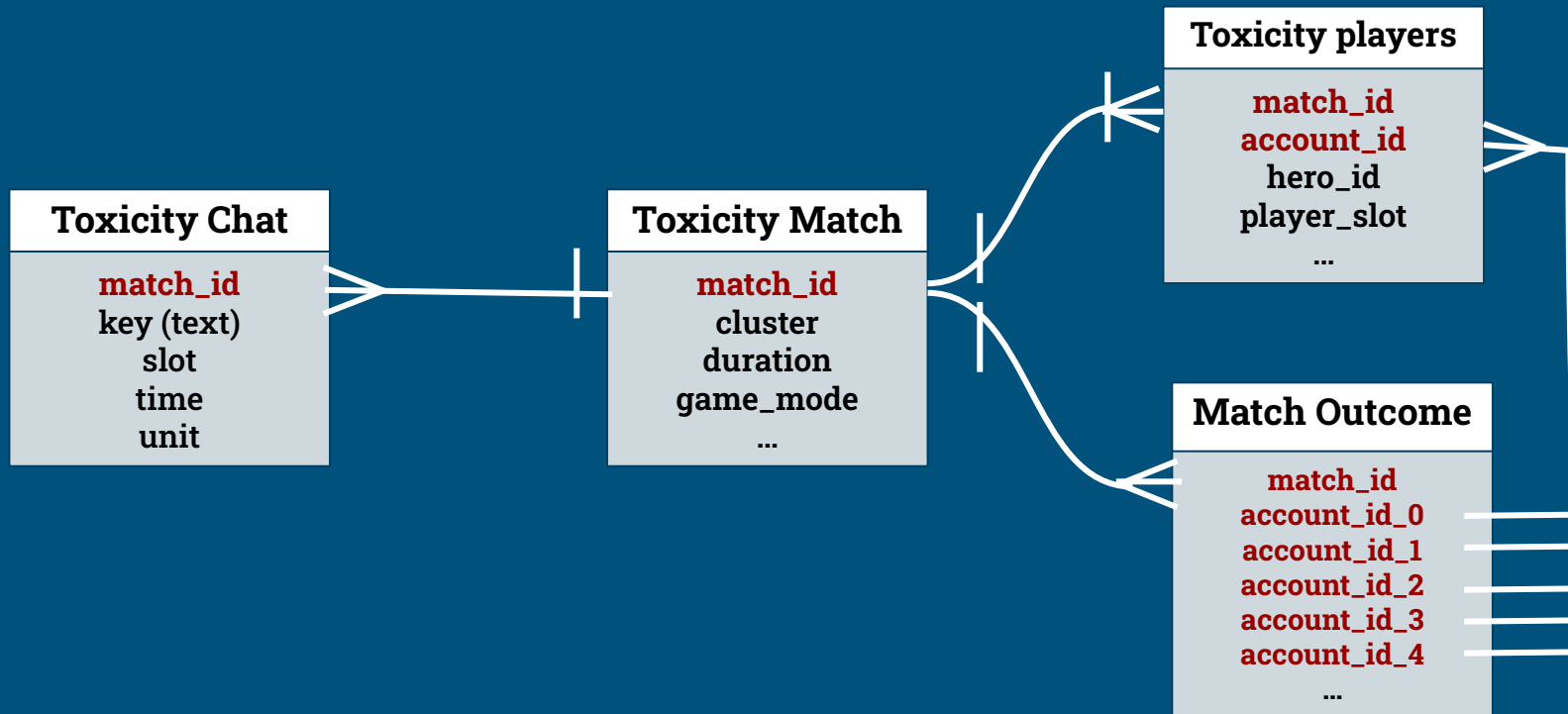
Visit Databricks Delta Lake



We leverage Databricks Delta Lake for **data ingestion**. We are able to build **Lakehouse architecture** after the importing the data from existing storage system (S3, HDFS)



Dota 2 Delta Lake Table Diagram



Build NLP Model with John Snow Lab

NLP Pipeline

Document
Assembler

Tokenizer
Normalizer

Stop Word
Lemma

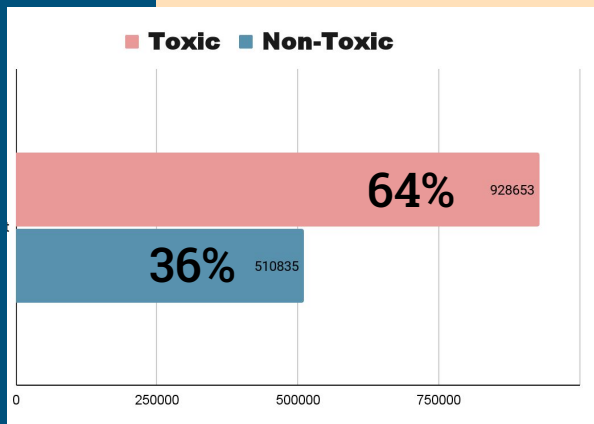
Word
Embedding

Sentence
Embedding

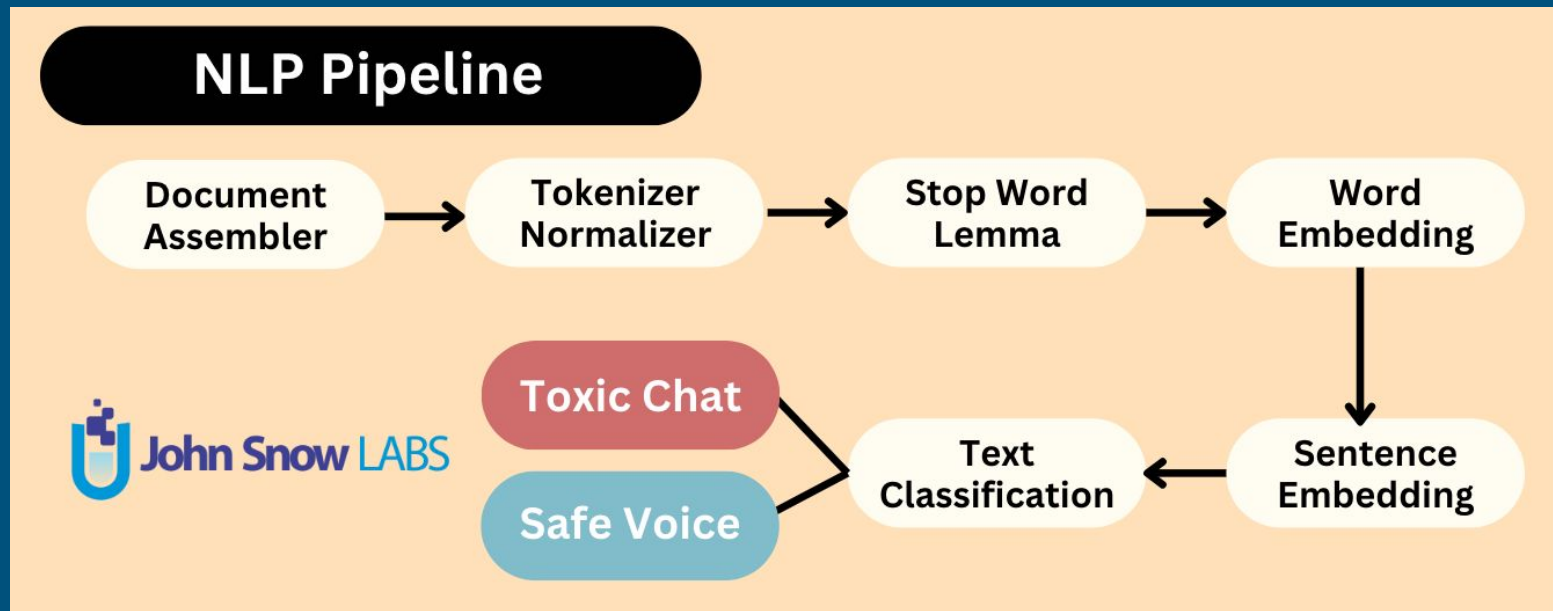
Text
Classification

Toxic Chat

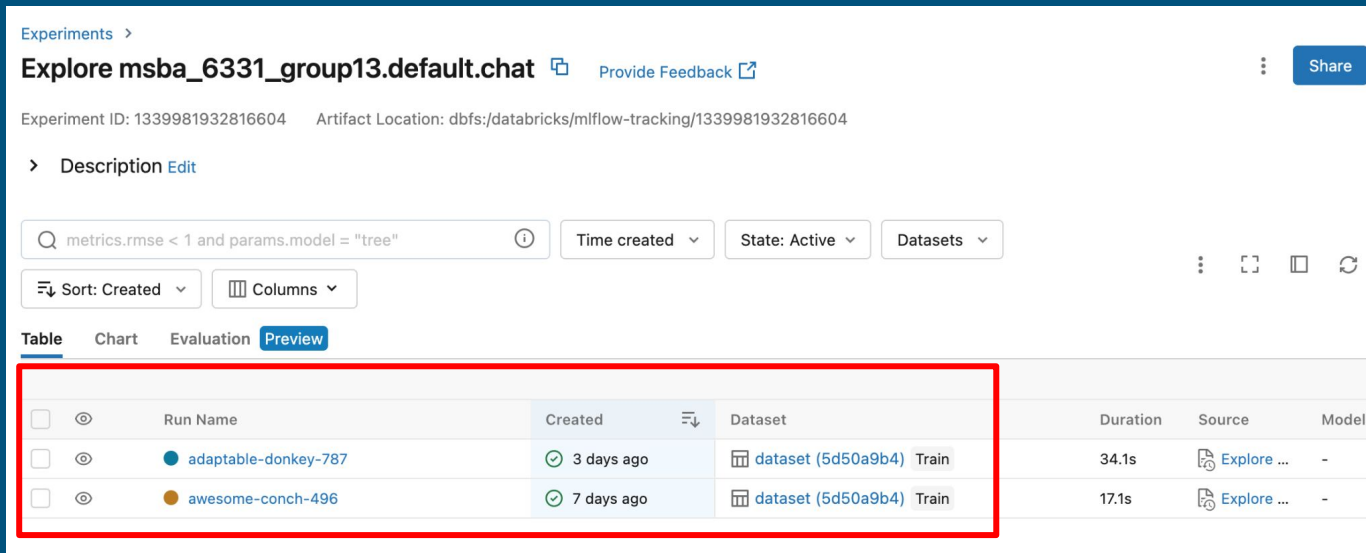
Safe Voice



Build NLP Model with John Snow Lab



Develop and Track Model with MLflow



The screenshot displays the MLflow Experiments interface. At the top, the experiment is titled "Explore msba_6331_group13.default.chat". Below the title, the Experiment ID is 1339981932816604 and the Artifact Location is dbfs:/databricks/mlflow-tracking/1339981932816604. A search bar contains the query "metrics.rmse < 1 and params.model = 'tree'". The interface includes filters for "Time created", "State: Active", and "Datasets". The "Table" tab is selected, showing a list of runs. A red box highlights the first two rows of the table, which are "adaptable-donkey-787" and "awesome-conch-496".

	Run Name	Created	Dataset	Duration	Source	Model
<input type="checkbox"/>	adaptable-donkey-787	3 days ago	dataset (5d50a9b4) Train	34.1s	Explore ...	-
<input type="checkbox"/>	awesome-conch-496	7 days ago	dataset (5d50a9b4) Train	17.1s	Explore ...	-

Run experiment with any ML library, framework or language, and automatically **keep track of parameters, metrics and code for experiments**

The Future of Toxicity Detection



Generative AI holds potential, despite the computational cost:

- Predict what players will say, before they say it
- Synthetic data
- Automatic labeling of training data
- Use of AI to label in real time, instead of a specific, trained model



Thank You!