Los Alamos NATIONAL LABORATORY

HPC Consult Ticket Analysis with SambaNova

HPC-SYS

Daisy Nsibu
Mentors: Mike Mason, Tolulope Olatunbosun

ROSY: 6a69c3df
LA-UR-25-28252

Background

- The HPC Division uses RT, a ticketing system that tracks user issues and consultant interactions
- RT contains 100k+ tickets spanning decades of user support history
- RT's poor search functionality and lack of analytical capabilities severely limit data utilization
- Consultants are unable to mine historical trends or identify common user issues

Objectives

- Provide a better method for searching tickets
- Develop a tool that uses AI/ML that provides comprehensive analysis of consult queue tickets.
- Use LLMs hosted by SambaNova to analyze the full ticket correspondence efficiently

Data Preprocessing

Data Acquisition

- Extracted over 100k consult queue tickets and corresponding child tickets from RT
- Extracted DST data from HPC Calendar

Data Transformation

- Normalization
- Categorical variable transformation
- Feature creation

Data Validation

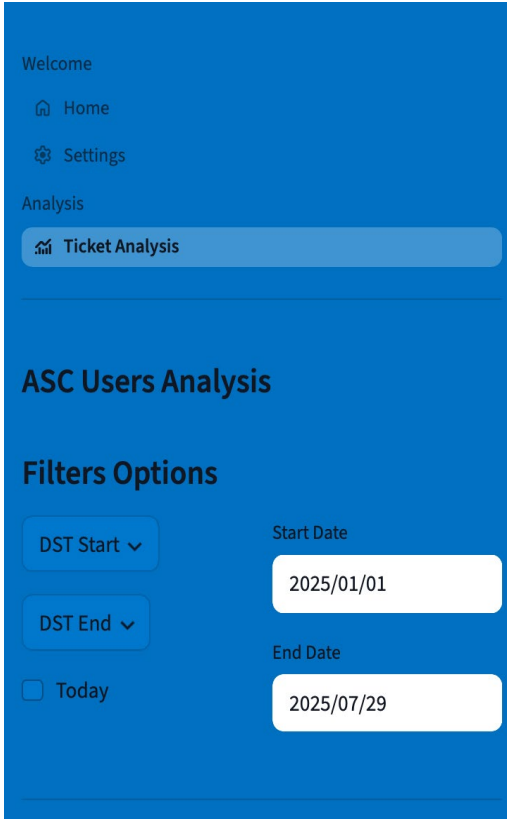
- Utilized Polars for data cleaning
- Handle null/missing values
- Remove duplicates
- Validate data types

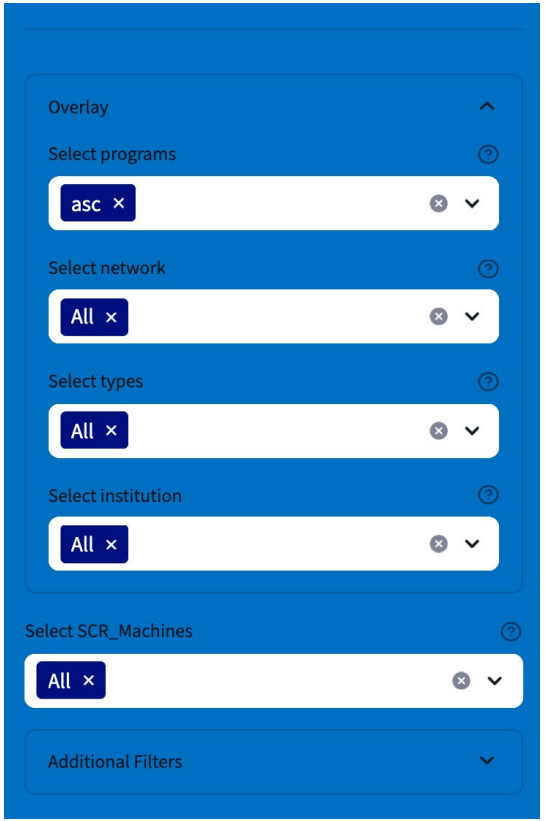
Data Reduction

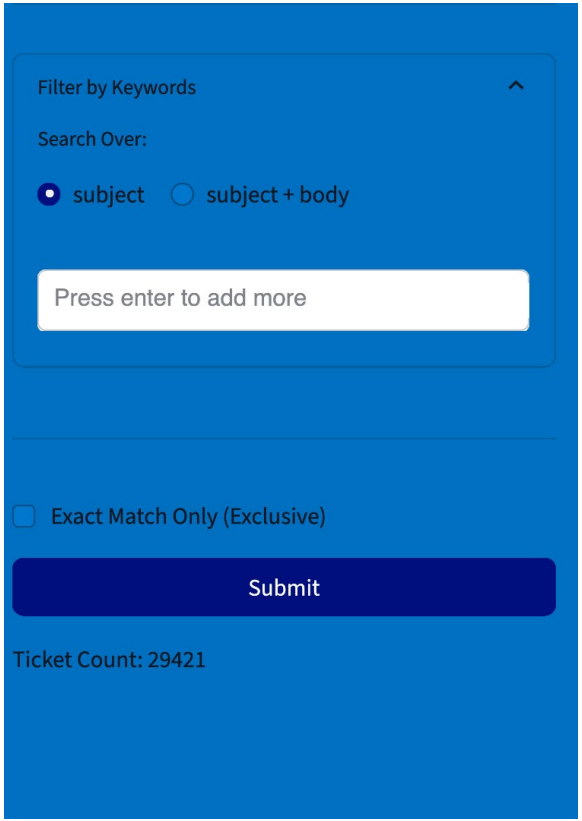
- Dimensionality Reduction
- Ticket correspondence aggregation

Features: Dynamic Filtering Capabilities

Features: Dynamic Filtering Capabilities

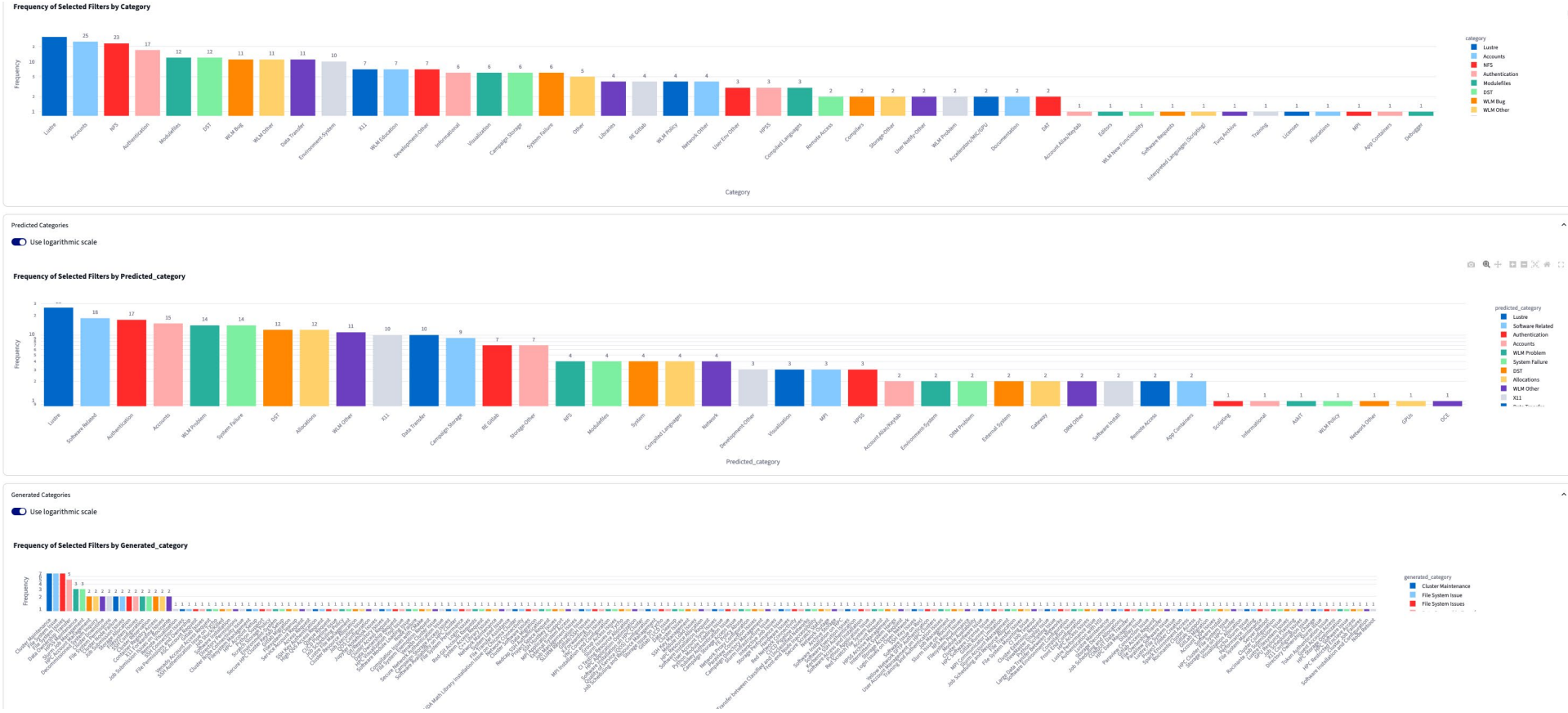






Los Alamos NATIONAL LABORATORY

Features: Summary View


Frequency of Selected Filters by Category
Frequency
Category
category
Lustre
Accounts
NFS
Authentication
Modulefiles
DST
WLM Bug
WLM Other
Data Transfer
Environment System
X11
WLM Education
Development Other
Informational
Visualization
Campaign Storage
System Failure
Other
Libraries
RE GitLab
WLM Policy
Network Other
User Env Other
HPSS
Compiled Languages
Remote Access
Compilers
Storage Other
User Notify Other
WLM Problem
Accelerators/HAC/GPU
Documentation
Dst
Account Alias/Keytab
Editors
WLM New Functionality
Software Requests
Interpreted Languages (Scripting)
Turq Archive
Training
Licenses
Allocations
MPI
App Containers
Debugger
Predicted Categories
Use logarithmic scale
Frequency of Selected Filters by Predicted_category
Frequency
Predicted_category
predicted_category
Lustre
Software Related
Authentication
Accounts
WLM Problem
System Failure
DST
Allocations
WLM Other
X11
Data Transfer
Campaign Storage
RE GitLab
Storage Other
NFS
Modulefiles
System
Compiled Languages
Network
Development Other
Visualization
MPI
HPSS
Account Alias/Keytab
Environment System
DRM Problem
External System
Gateway
DRM Other
Software Install
Remote Access
App Containers
Scripting
Informational
KASIT
WLM Policy
Network Other
GPUs
OCE
Generated Categories
Use logarithmic scale
Frequency of Selected Filters by Generated_category
Frequency
generated_category
Cluster Maintenance
File System Issue
File System Issues

What are the Top Issues Users are Facing?

Total Tickets Analyzed: 245

1. Account Access Issues (32 tickets, 13.0%)

User Question:

Users frequently report being unable to log into their accounts on [specific clusters]. Common symptoms include authentication failures, expired credentials, or permission issues.

Resolution:

The standard resolution involves verifying account status in the user database, resetting credentials if necessary, and ensuring proper group permissions are set.

2. Project Space and Quota Issues (23 tickets, 9.0%)

User Question:

Users request project space, inquire about quota limits, or report issues with project space.

Resolution:

The resolution involves creating project space, setting quota limits, or resolving issues with project space.

3. Job Submission and Scheduling Issues (20 tickets, 8.0%)

User Question:

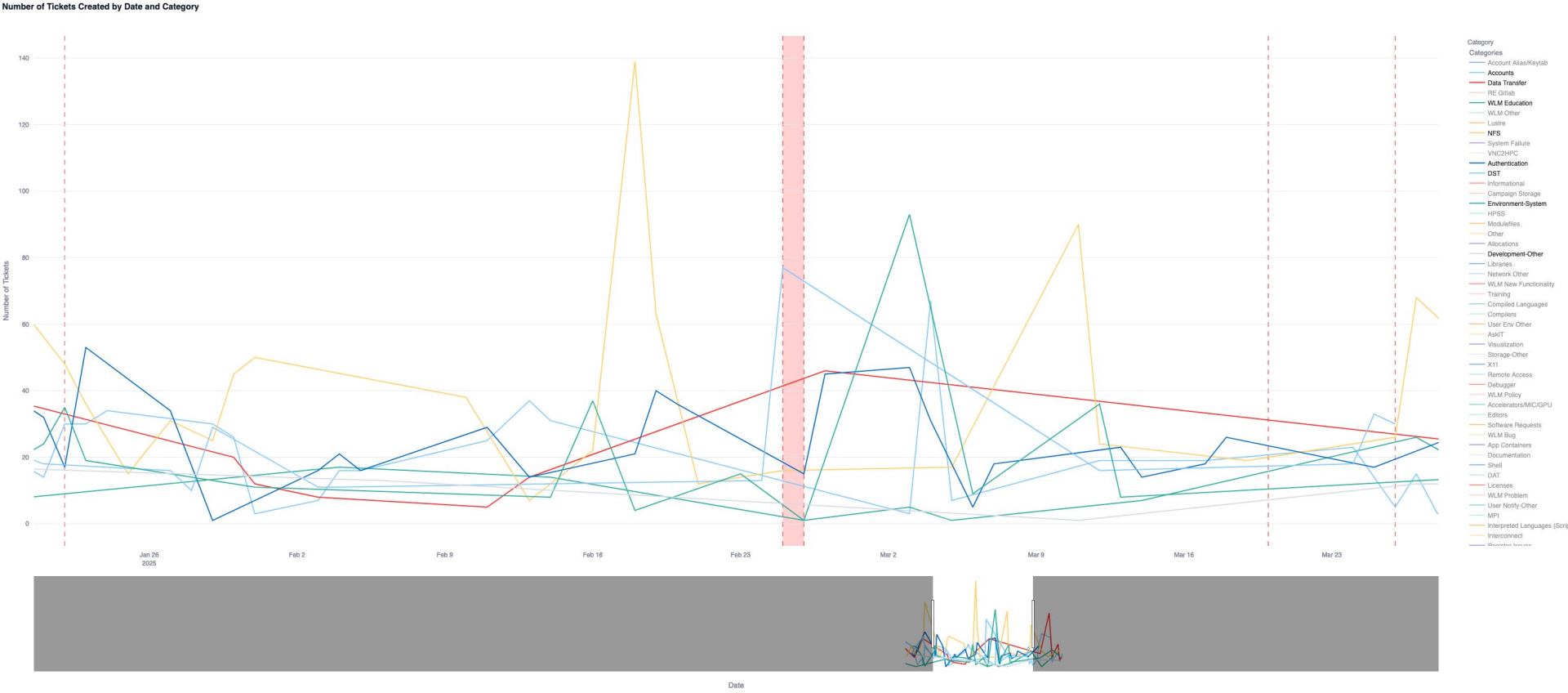
Users report issues with job submission, job scheduling, or job priority.

Resolution:

The resolution involves troubleshooting job submission issues, adjusting job scheduling policies, or resolving job priority issues.

Features: Time Analysis

Features: Time Analysis with DST Overlay

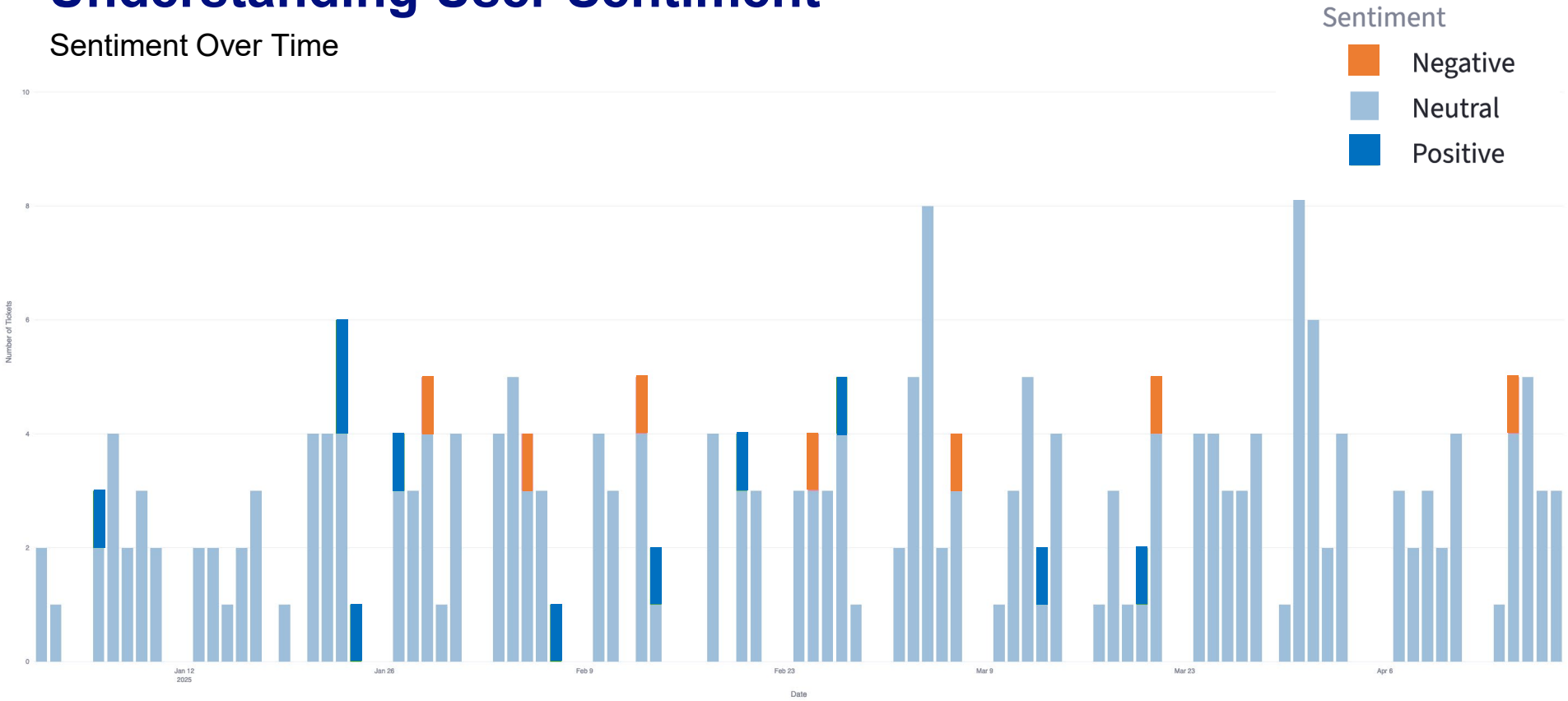


Los Alamos NATIONAL LABORATORY

Features: Ticket Activity

Understanding User Sentiment

Sentiment Over Time



Los Alamos NATIONAL LABORATORY

Features: Anomaly Detection

Features: Anomaly Detection

Ticket Detective

Find Anomalies

Analysis complete! Found 20 anomalies.

Anomaly Detected:**Unexpected spikes in ticket volume:** There is a noticeable spike in ticket volume on March 21st (Tickets 261326, 261332, 261335, and 261356), which could indicate an underlying issue or a coordinated effort to submit tickets.

Anomaly Detected:**Issues that are normal in one context but anomalous in another:** Ticket ID 260537, where a user is running a code that uses up all 2048 MPI communicators, is normal in the context of the user's work but anomalous in the context of the system's configuration.

Anomaly Detected:**Tickets that seem routine but have unusual details in their description:** Ticket ID 258060, where a user is experiencing an error when trying to migrate files from /lustre/xrscratch1/ to /campaign/ storage using conduit on redcap, seems routine but has unusual details in its description, such as the error message 'Failed to create grpc client: failed to dial into conduit server: context deadline exceeded'.

Anomaly Detected:**Groups of tickets that together represent unusual behavior:** Tickets 258602, 259078, and 261326, which all relate to issues with scratch5 or scratch spaces, could indicate a larger issue with the scratch filesystem or the way users are interacting with it.

Anomaly Detected:**Sequential tickets that form a suspicious pattern:** Tickets 258048, 258122, and 258149, which all relate to issues with Tycho front-end nodes, could indicate a pattern of issues with the Tycho cluster or its front-end nodes.

Sequential tickets that form a suspicious pattern: Tickets 258048, 258122, and 258149, which all relate to issues with Tycho front-end nodes, could indicate a pattern of issues with the Tycho cluster or its front-end nodes.

Potential Impact on System Reliability or Security:

Reccomendations:

Impact:

on System Reliability or Security:**

Tickets:

Sequential tickets that form a suspicious pattern: Tickets 258048, 258122, and 258149, which all relate to issues with Tycho front-end nodes, could indicate a pattern of issues with the Tycho cluster or its front-end nodes.

Features: Comparative Analysis

Comparative Analysis View

Stop

← Return to Main View

IC Machines Analysis

Summary | Time Analysis | Ticket Activity | Anomaly Detective | Data Gallery

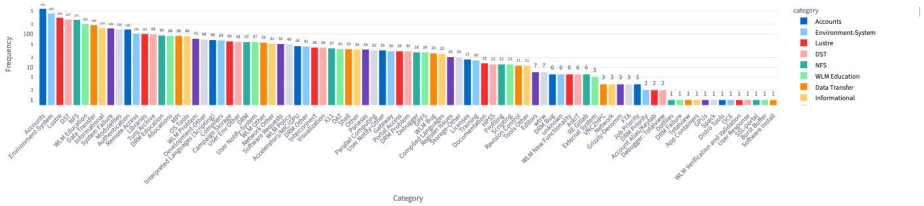
IC Machines Analysis

Filtered Data

Categories

Use logarithmic scale

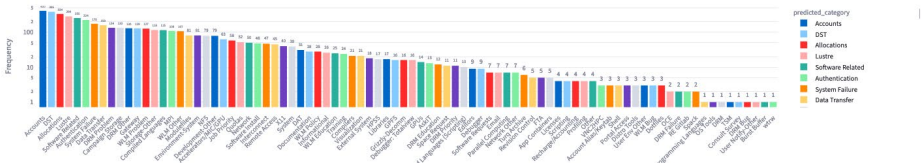
Frequency of Selected Filters by Category



Predicted Categories

Use logarithmic scale

Frequency of Selected Filters by Predicted_category



ASC Machines Analysis

Summary | Time Analysis | Ticket Activity | Anomaly Detective | Data Gallery

Time Analysis

Select a filter:

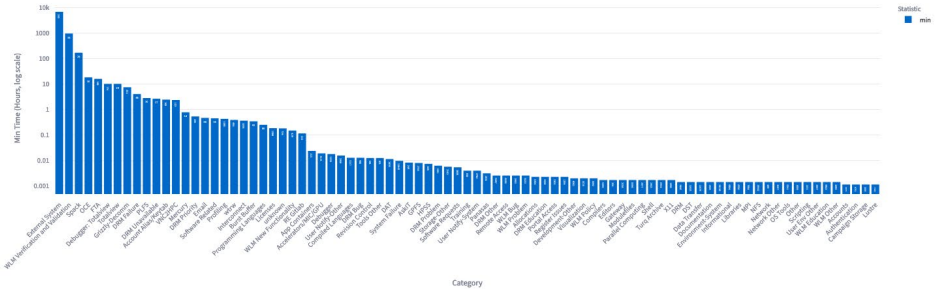
started | created

Select a statistic:

min | max | mean | median

Use logarithmic scale

Min Resolution Time by Category



Frequency Table

Summary Table

Number of Tickets Created & Resolved by Date

Select Clusters for DST Dates

Choose an option

Show Control Chart

Los Alamos NATIONAL LABORATORY

Challenges

Challenges: Processing Data

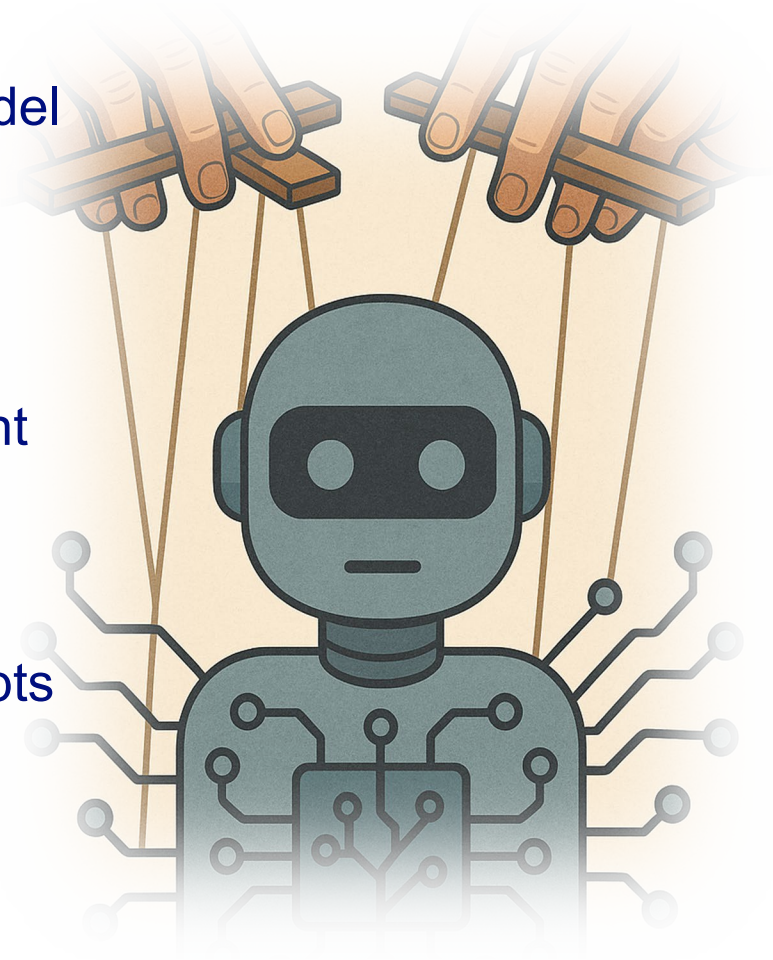
- Real-Time Processing
  - Overloaded SN platform by sending the full dataset at once
  - Instant processing strained compute resources

- Batch Processing
  - Split workload into “smaller” batches to reduce resource strain
  - Still resource-intensive — caused SN overload
  - Required off-peak scheduling (6:30 PM–6:30 AM) when few users are active to avoid disruption and optimize compute resources

Challenges: Prompt Engineering

- Developed 20+ prompts to instruct the model to:
  - Summarize tickets and resolutions
  - Predict and generate HPC relevant categories
  - Justify predictions and detect sentiment
  - Uncover user's main issues
  - Find Anomalies in data

- Challenge: Batch processing limited prompts to batches and not the full data set — requiring manual result aggregation



Challenges: Hallucinations

- Model generated categories that were not in the list of consultant defined categories
- Sentiment outputs deviated from expected labels (Positive, Neutral, Negative)
- Required post-processing to detect and remove hallucinated content, followed by reprocessing the affected data

Conclusion

Summary

- Developed the first of its kind, AI powered web application, that consultants can use to easily explore and analyze ticket data

- User friendly platform: User experience is prioritized, and consultants are kept updated all the time

Impact

- Consultants have insights into ticket activity, trends over time, general system issues.

- Efficiency gains would free consultants to tackle more complex technical challenges

Future Work

❑ Use More Advanced LLMs

- Process Ticket data using Llama 4 maverick

❑ Develop Hallucination Guardrails

- Use LLM to detect hallucinations

❑ Advanced Prompt Engineering Techniques

- Use AI to help create better prompts

❑ Live Ticket Integration

- The web app connects to RT to analyze tickets in real-time

Thank You!

- Mentors: Mike Mason , Tolulope Olatunbosun

- A Special Thanks to Hunter & Anna From HPC-ENV

Questions ?