

ACCESS TO TECHNOLOGY



Knoxicle: AccessGuru - Accessibility Insights

Analyzing Web Accessibility Violations (WCAG 2.1)

Total Violations

3524

Unique Pages

591

Critical Errors

475

Avg Severity

3.55

By Knoxicle (Kannika Armstrong)

DUBSTECH Information School UNIVERSITY OF WASHINGTON APPLIED ANALYTICS
DATATHQN
the 7th data hackathon

Machine Learning & Risk Modeling

Model Accuracy

97.4%

Margin of Error

±0.3%

How the Model Works

Predict Violation Impact

Select attributes to estimate how severe a web accessibility violation will be.

Industry/Domain

e-commerce

Violation Category

Layout

Specific Violation

avoid-inline-spacing

Run Model Prediction

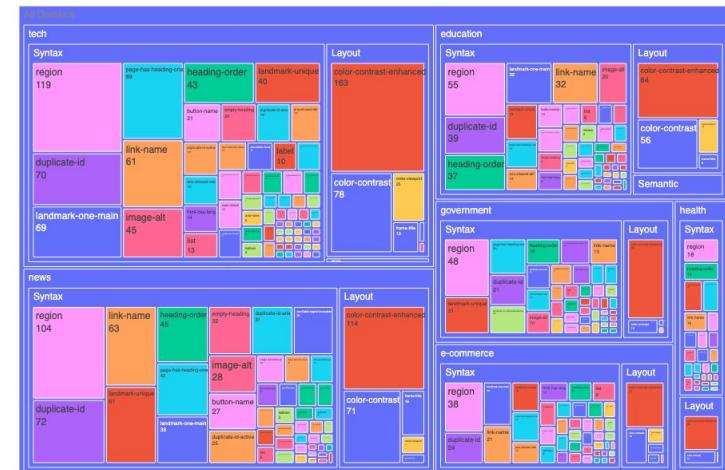
Prediction Result: SERIOUS

Analysis Summary:

- Domain: e-commerce
- Violation Category: Layout
- Violation Name: avoid-inline-spacing

The AI model determined that this specific configuration typically results in a serious impact on digital accessibility.

Hierarchical View of Accessibility Failures



Projects Agendas

- Data Analytics and Data Visualization
- Machine Learning / Predictive Modeling

Live Demo: <https://accessguru-knoxicle.streamlit.app/>

GitHub Repo: <https://github.com/A-Kannika/DubsTech-Datathon-2026-Knoxicle>



Knoxicle: AccessGuru - Accessibility Insights

Answering questions: Data Analytics and Data Visualization (1)

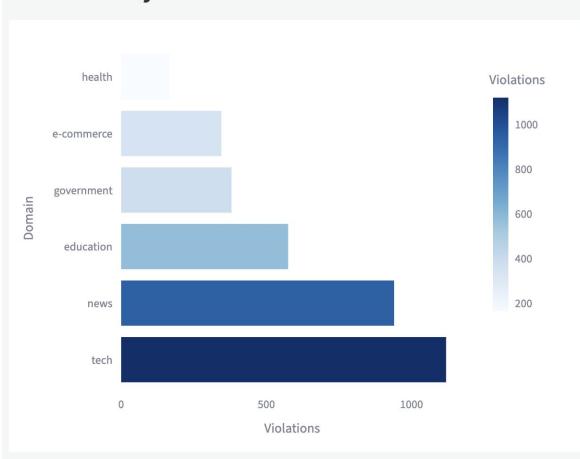
Q: Which domains have the highest number of accessibility violations?

A: Tech domain.

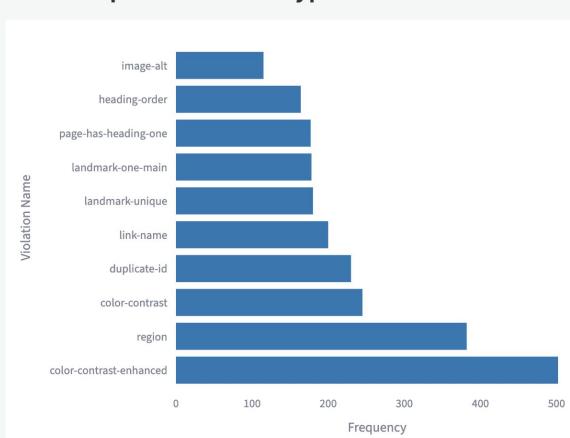
Q: What violation types are most common across sites or domains?
A: Syntax.

Q: Are there patterns in violations by violation category?
A: Yes.

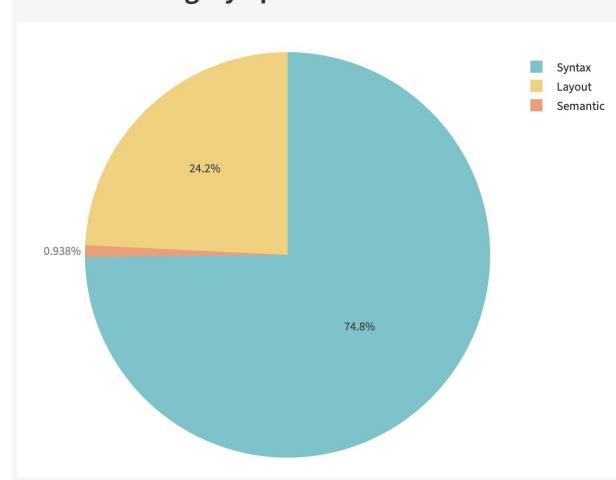
Violations by Domain



Most Frequent Violation Types



Violation Category Split



Answering questions: Data Analytics and Data Visualization (2)

Q: Identify websites/pages with the most severe accessibility issues.

A: <https://www.pluralsight.com/>.

Top 10 Inaccessible Pages

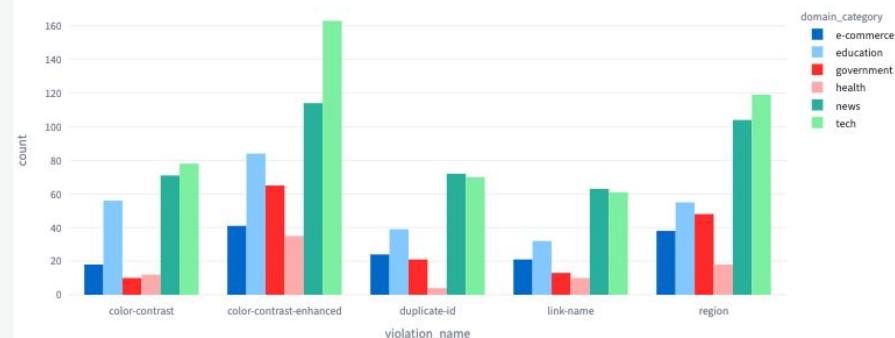
Rank	web_URL	total_severity	error_count
1	https://www.pluralsight.com	258	69
2	https://www.coursera.org	153	45
3	https://www.edx.org	153	39
4	https://arstechnica.com/health/	124	36
5	https://arstechnica.com/science/	124	36
6	https://arstechnica.com/gadgets/	124	36
7	https://www.geeksforgeeks.org	96	28
8	https://www.spss.com	70	18
9	https://www.cloudacademy.com	66	17
10	https://www.tampabay.com	58	16

Q: Examine invisible barriers: e.g., are certain violations more common on government sites than on e-commerce sites?

Compare Invisible Barriers

Pick domains to compare side-by-side:

government x news x e-commerce x tech x health x education x



Answering questions: Data Analytics and Data Visualization (3)

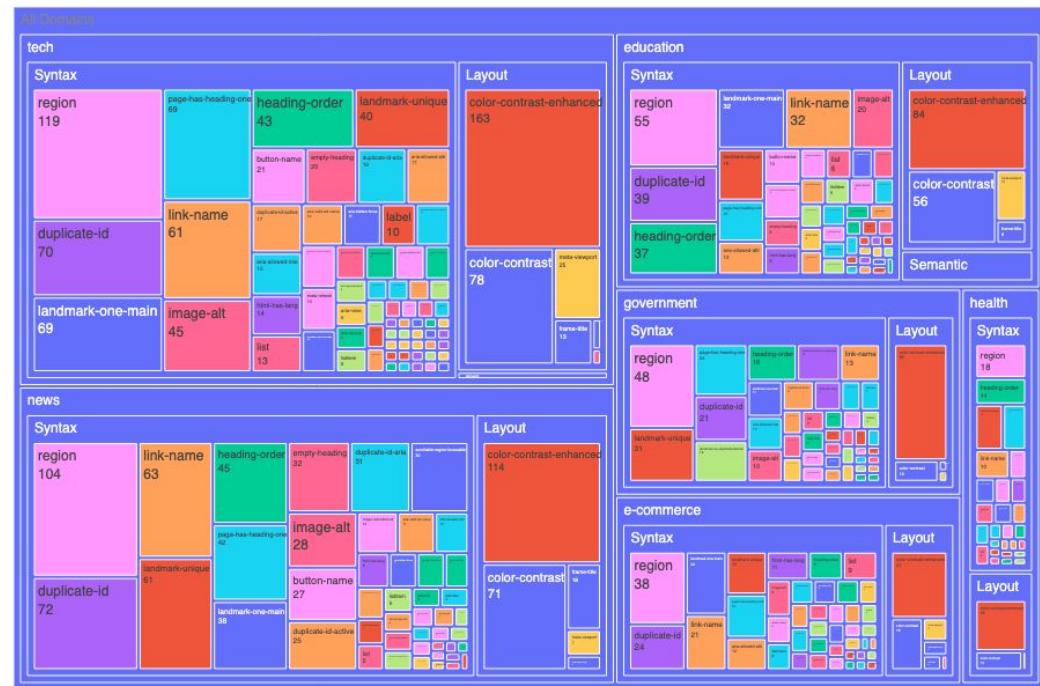
Q: Create visualizations showing where technology fails: heatmaps of violations, bar charts by domain, or trend analyses.

Violation Density Heatmap

Heatmap: Domain vs. Violation Category



Hierarchical View of Accessibility Failures



Machine Learning / Predictive Modeling (1)

Model Performance & Validation

To ensure the reliability of our accessibility risk predictions, we validated the model using an 80/20 train-test split:

- Algorithm: Random Forest Classifier (100 Estimators)
- Model Accuracy: 97.4% (Calculated via Mean Accuracy on unseen test data)
- Margin of Error: $\pm 0.3\%$
- Validation Method: Hold-out validation to prevent overfitting and ensure the model generalizes well to new, unseen websites.

Machine Learning & Risk Modeling

Model Accuracy

97.4%

Margin of Error

$\pm 0.3\%$

> How the Model Works

Predict Violation Impact

Select attributes to estimate how severe a web accessibility violation will be.

Industry/Domain

e-commerce

Violation Category

Layout

Specific Violation

avoid-inline-spacing

Run Model Prediction

Prediction Result: SERIOUS

Analysis Summary:

- Domain: e-commerce
- Violation Category: Layout
- Violation Name: avoid-inline-spacing

The AI model determined that this specific configuration typically results in a **serious** impact on digital accessibility.



Knoxicle: AccessGuru - Accessibility Insights

Machine Learning / Predictive Modeling (2)

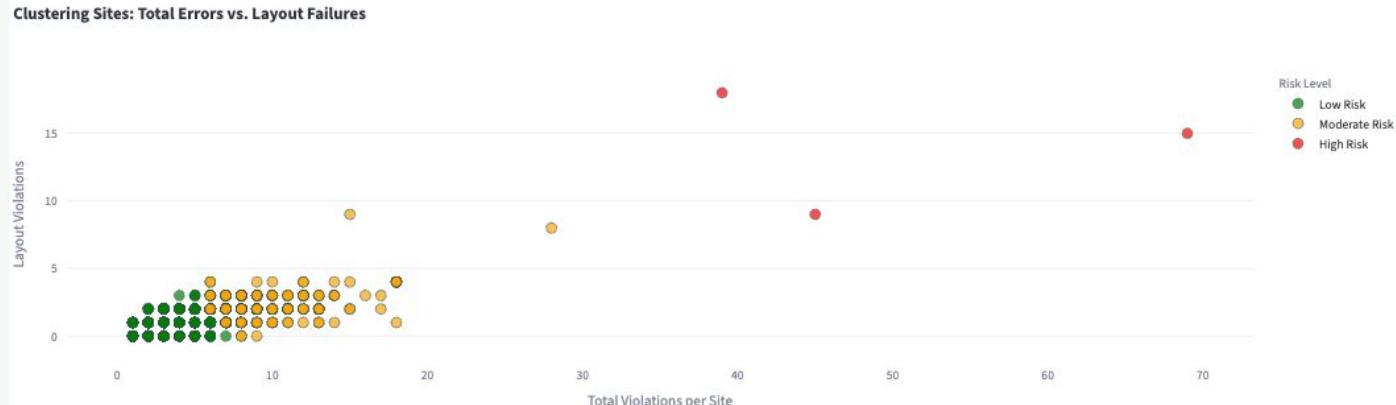
Q: Cluster websites based on similarity in violation patterns to identify “high-risk” domains.

Website Risk Clustering

We used K-Means Clustering to group the 448 websites by their 'failure patterns'.

Select Violation Category for Y-Axis Analysis:

Layout



Insight: By switching to Layout, you can see if the 'High Risk' cluster is driven specifically by that type of violation or if it's a general spread across all categories.



Machine Learning / Predictive Modeling (3)

Ranking of Inaccessible Design

Rank domains based on the likelihood of encountering specific impact levels.

Select Impact Level to Rank By:

critical

Domain	Likelihood of Critical Issues	
tech	<div style="width: 16.26%; background-color: red;"></div>	16.26%
education	<div style="width: 15.30%; background-color: red;"></div>	15.30%
e-commerce	<div style="width: 12.46%; background-color: red;"></div>	12.46%
news	<div style="width: 12.23%; background-color: red;"></div>	12.23%
health	<div style="width: 10.30%; background-color: red;"></div>	10.30%
government	<div style="width: 7.89%; background-color: red;"></div>	7.89%

Q: Rank domains or pages by likelihood of inaccessible design.

Risk Summary: Critical Impact

Key Findings:

- **Highest Risk Domain:** Tech
- **Concentration:** 16.3% of violations in this domain are critical.

Digital Equity Insight: When a domain has a high concentration of **critical** issues, it indicates a systemic failure in the design process of that industry. For users, this means the barrier isn't just a one-off mistake, but a pattern that makes these types of sites (like Tech) fundamentally harder to access.