



OPEN SOURCE MACHINE DATA HUB FOR PROGNOSTICS AND HEALTH MANAGEMENT

VISIT [HTTPS://MACHINEDATAHUB.AI/](https://machinedatahub.ai/)

STUDENTS: CECILIA BARNES, JIWOO PARK, MATTHEW SULLIVAN



Motivation

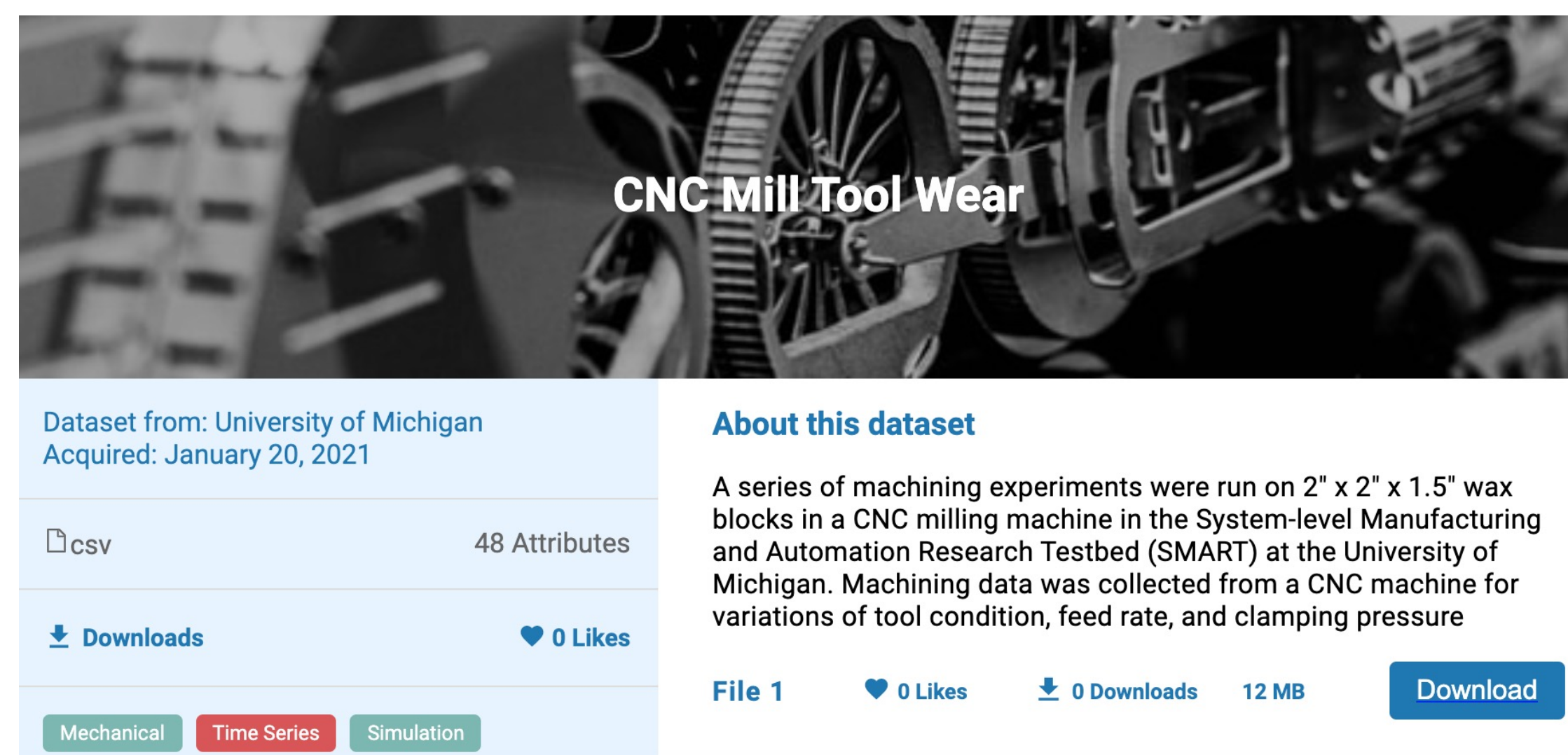
The goal of this project is to make state of the art machine learning and artificial intelligence techniques more accessible to engineering domains by providing suitable benchmark datasets for advancing prognostics.

Goals

1. Deploy a fully functioning, open-source front-end web app to view and download reliable prognostics and health management datasets that includes accurate metadata for all datasets
2. Publish an open-source Python package for users to view and download datasets from their command line
3. Create machine learning use case examples for users to view on blog

Data Curation

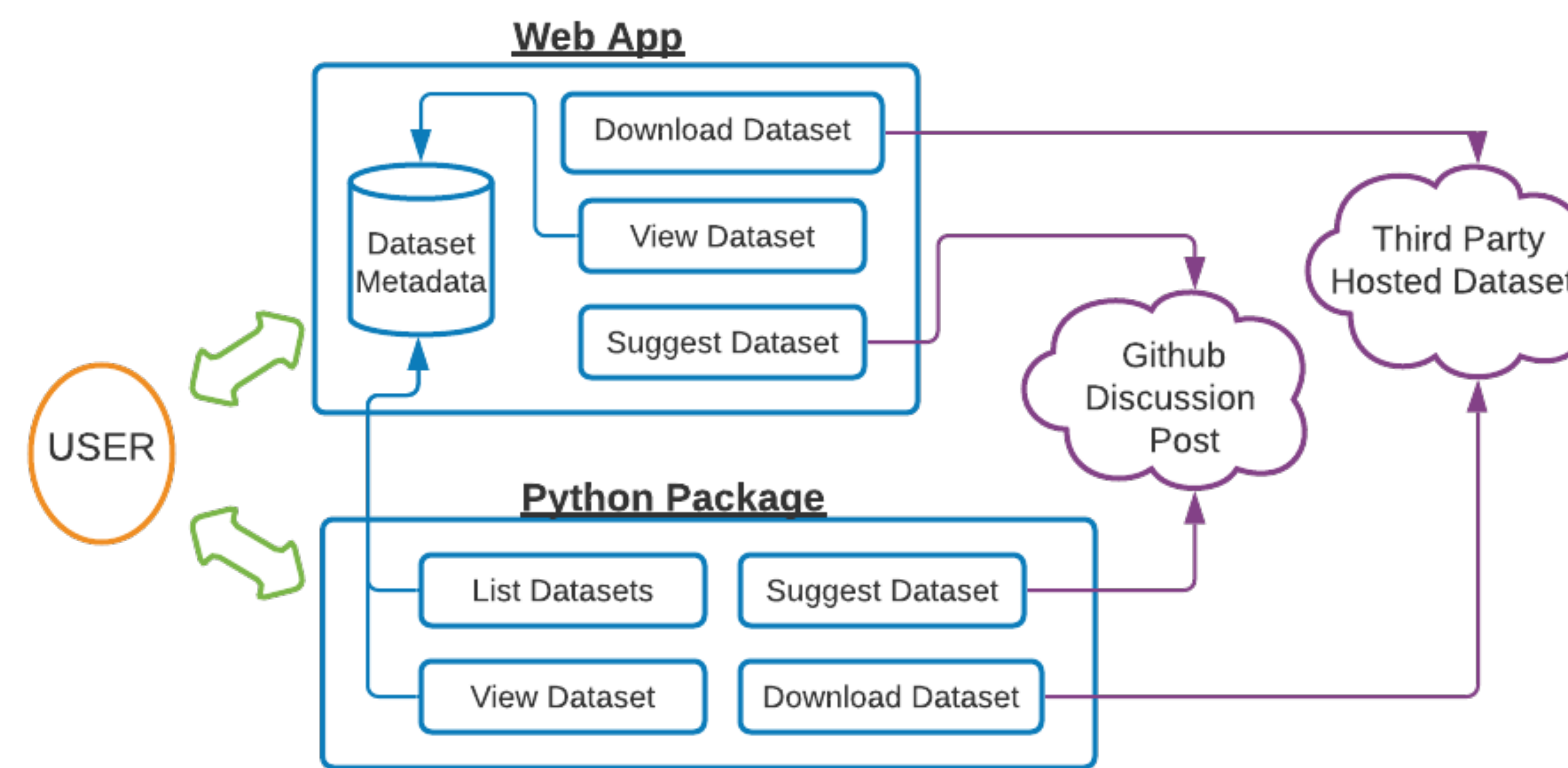
- Datasets were acquired from reliable sources such as research institutions and engineering companies
- Each dataset is inspected for accuracy, integrity, and readability, and detailed metadata gathered for each dataset so users can sort and filter efficiently



Machine Data Hub Python Package

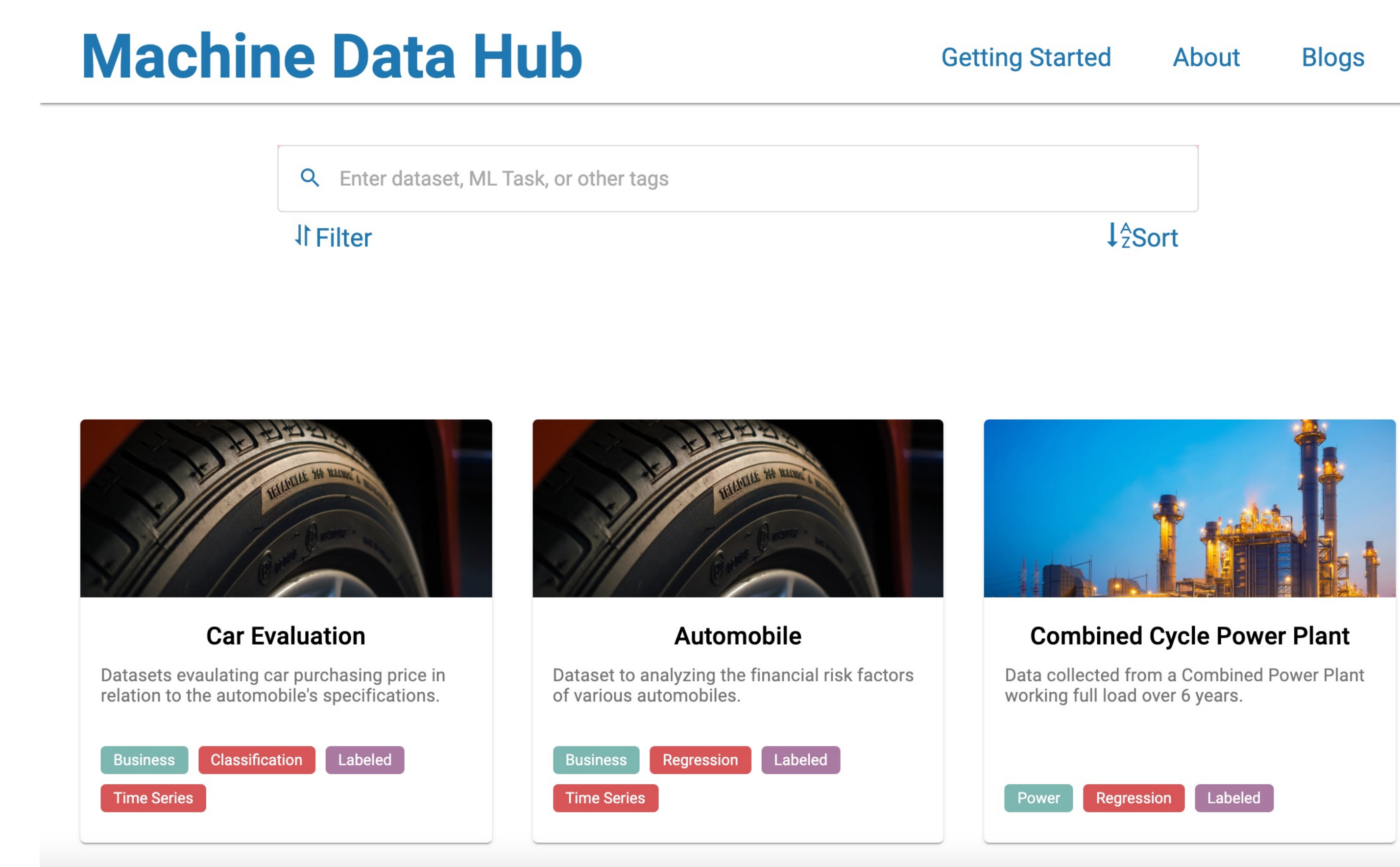
- Users can use `pip install machine-data-hub` to begin using our product from any operating environment

Project Architecture



Machine Data Hub Website

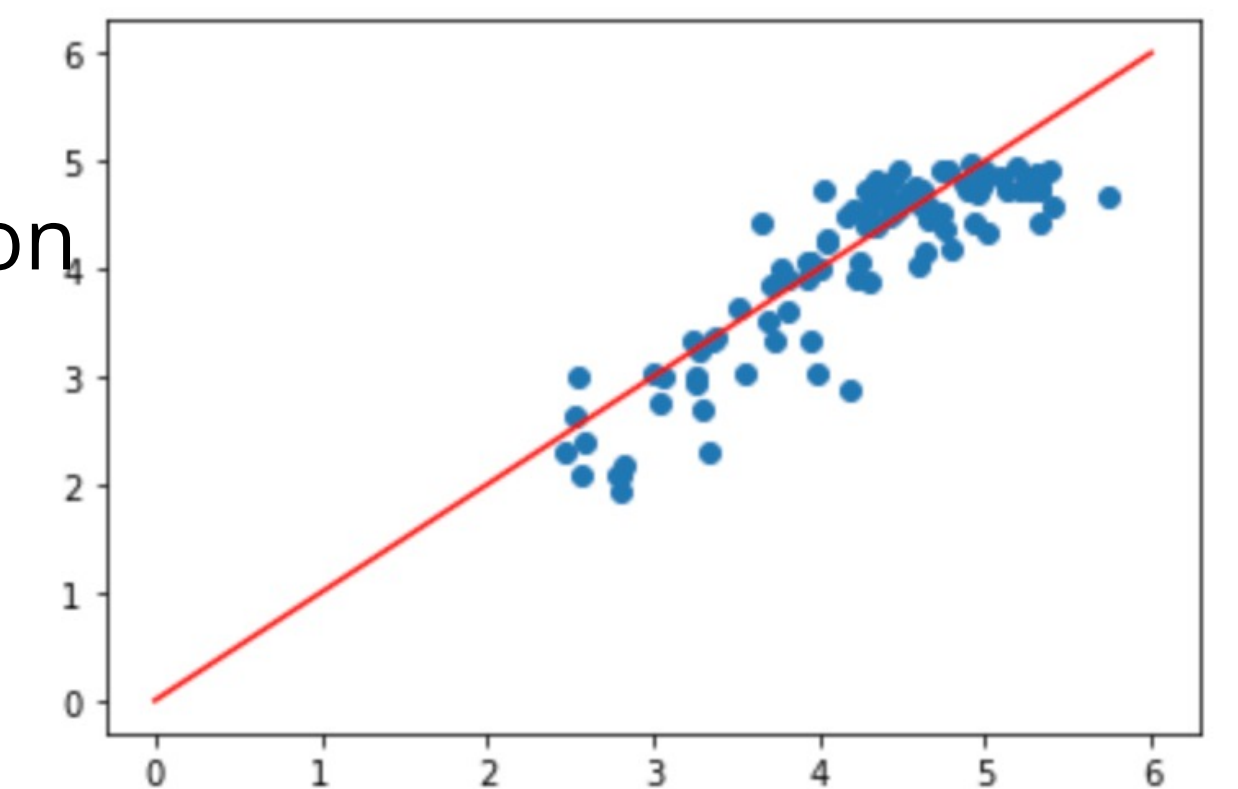
- Score of 93 from Lighthouse for performance and usability
- Received extensive feedback from UI/UX experts in industry and academia to make sure the site is user-friendly
- Similar datasets are recommended based on tags



Use Cases of the Machine Data Hub

Remaining Useful Life Prediction

- Used NASA's Turbofan Engine Degradation Simulation Dataset to predict how many cycles were left in the engine
- The figure to the right shows true remaining useful life on the x-axis vs the predicted remaining useful life in cycles



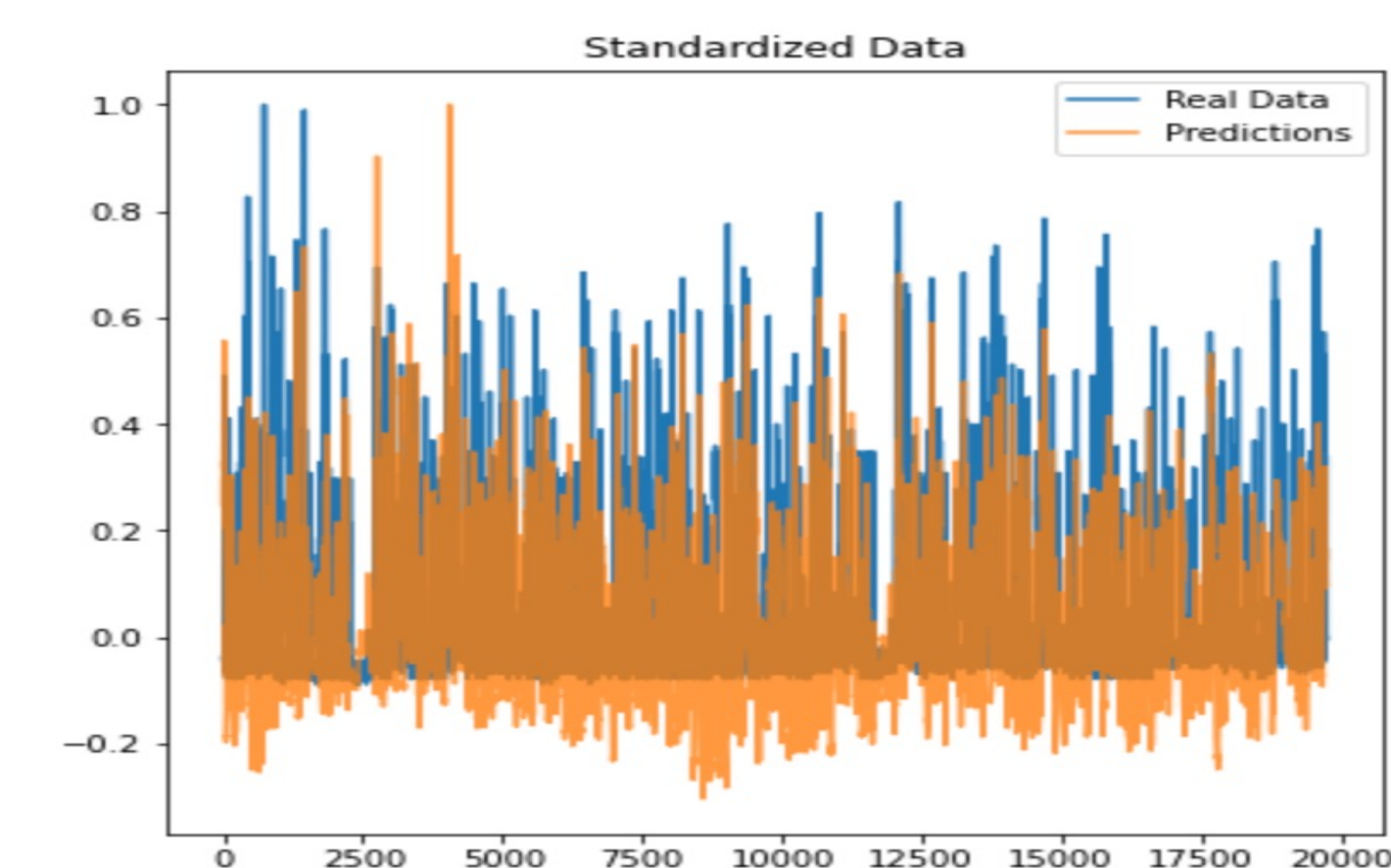
Energy Output Prediction

- Used UC Irvine's Combined Cycle Power Plant dataset to predict the energy output of the plant
- Found that a Random Forest Regressor using temperature, ambient pressure, and exhaust vacuum as inputs, or *df_3* in the results table to the right, was most accurate

	df_1	df_2	df_3	df_4
RMSE	5.677	3.501	3.273	3.203
R ²	.889	.9579	.963	.965
MAE	4.4365	2.6216	2.39	2.35

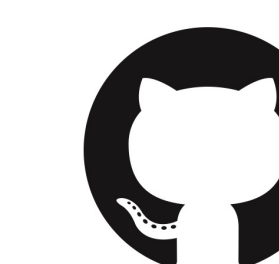
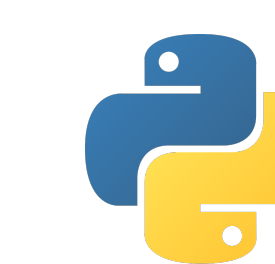
Energy Usage Prediction

- Used UC Irvine's Appliances Energy dataset to predict the energy output of the plant
- Created Linear LASSO model that predicted energy usage in houses in Belgium given temperature



Future Work and Acknowledgements

- Partner with UW Data Science and engineering courses to encourage more users and feedback
- Implement ETL process in Python package to convert all filetypes to .CSV
- Auth-N (Authentication) implemented
- Website analytics (page views, dataset downloads, user likes) with Postgres



HUMAN CENTERED DESIGN & ENGINEERING
UNIVERSITY of WASHINGTON

