

Efficient Training Data Label Collection for End-of-Life Fabric Sorting

Completed in collaboration with:

refiberd.

MIMS 2024 Capstone

Team Members: Abdullah Azhar, Mustafa Hameed, Erin Jones, Isidora Rollán, Prashant Sharma

Advisor: Dr. John Chuang

Information Technology for Recycling Textiles



Used clothing dump at Atacama Desert, Chile (11 - 59k tons of fabric)

Fashion Industry = 10%+ of global carbon emissions – more than international flights and maritime shipping combined.¹

Less than 15% of used clothing is donated and only 0.5% is recycled²

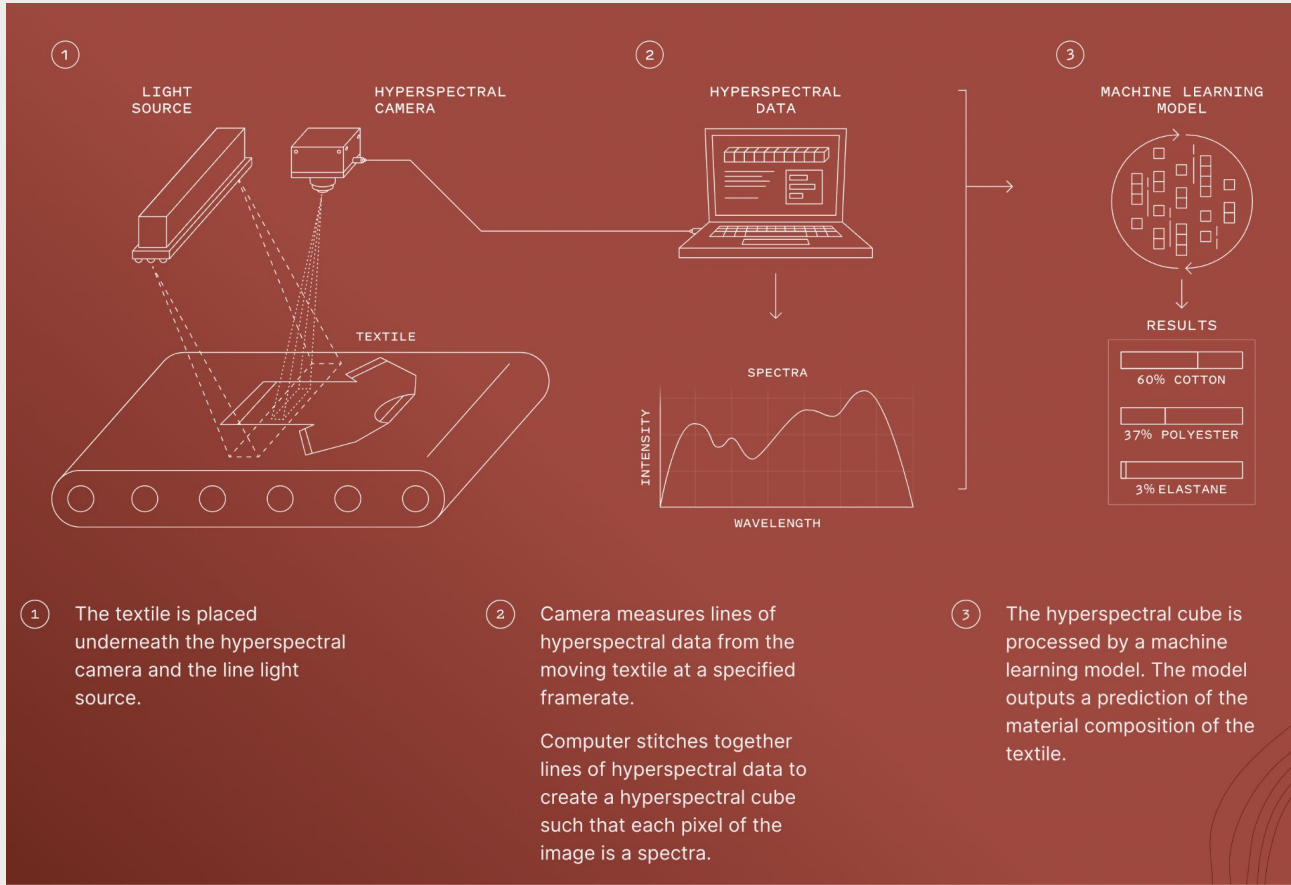
Big Problem - Too big for a 6 month capstone...

1. European Parliament

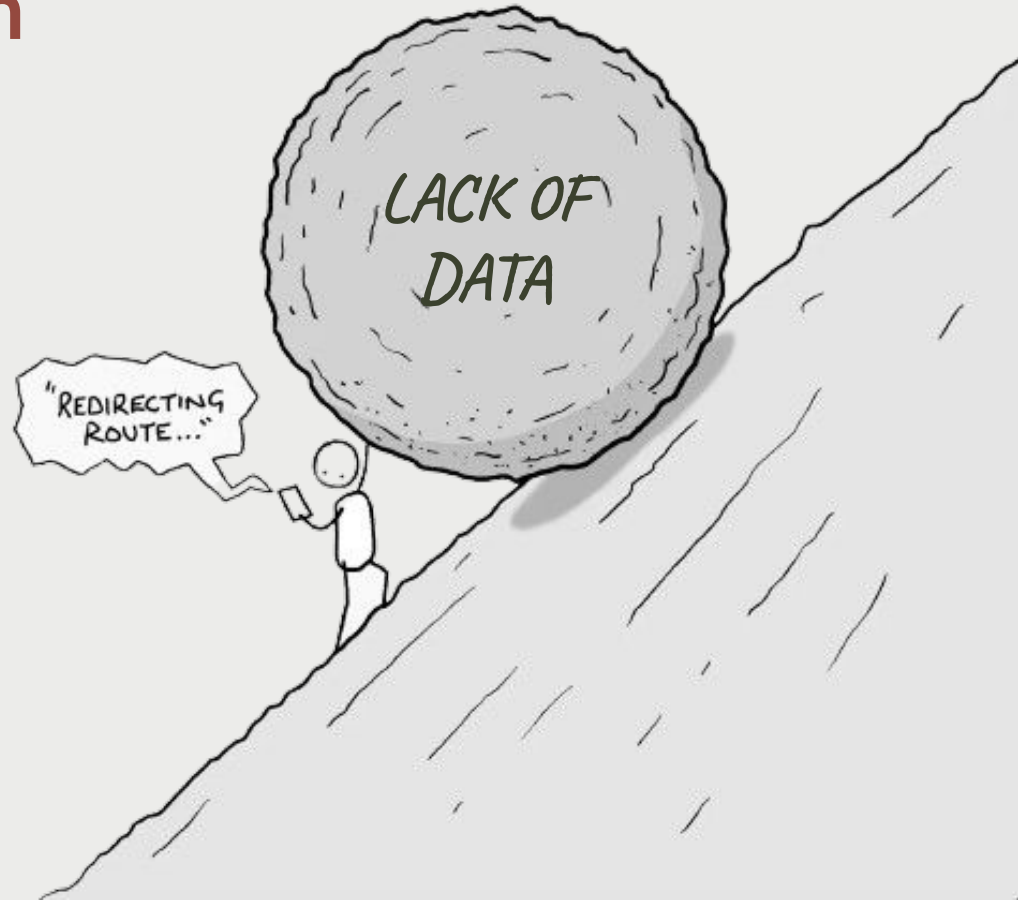
2. National Institute of Standards Technology

Who is Refiberd?

‘Advanced material detection via supervised regression + AI-based hyperspectral imaging’



The Barrier to Applying ML/AI in Climate



The Problem

Currently, labeling training data is a manual process



Manual Entry on
Google Sheets

2 minutes

Average Time of
**Entry with 2
people**

20,000

Target Samples
for ML model by
July

A bottleneck that Refiberd has is the generation of new data points



The Solution

They asked us to create a tool to automate the labeling process

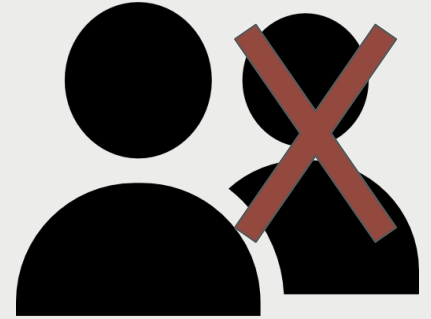
Constraints



Using AWS environment

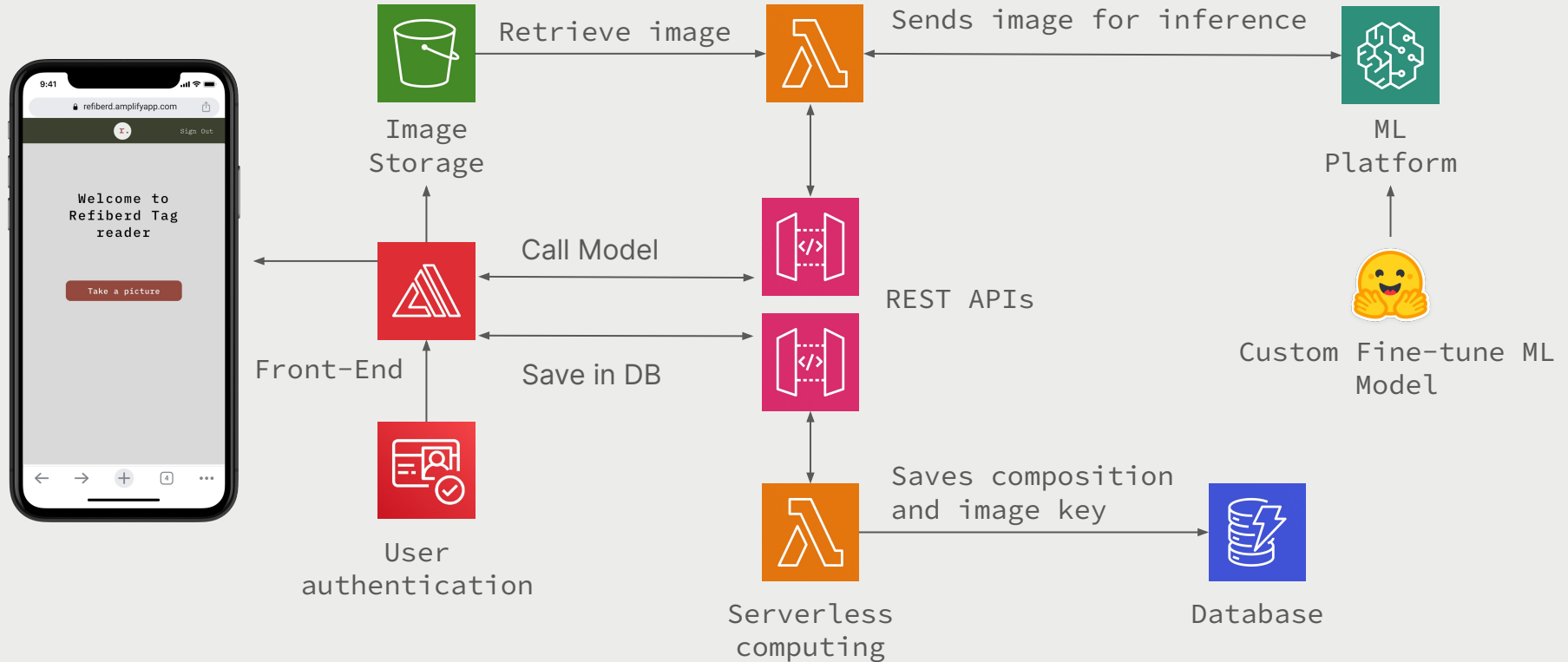


Mindful in costs



One person job

We leverage AWS architecture to create a robust tool



The model was essential to our solution

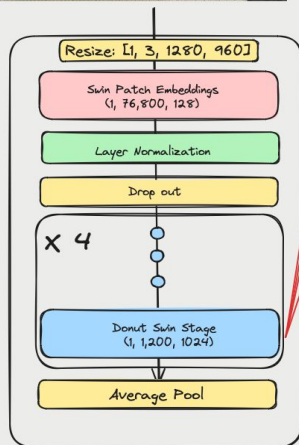
OCR Free Document Understanding Transformer (DONUT) - 220 M parameters

Fine-tuned on a 469 image-to-text training dataset

Cross-Entropy Loss for next-token prediction

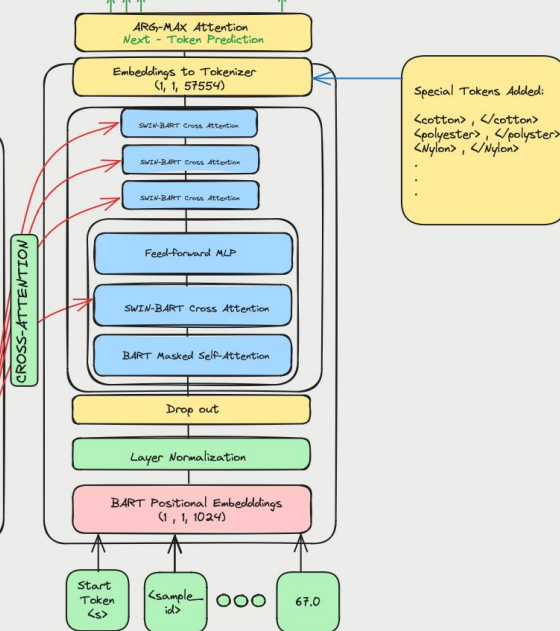
Levenshtein Distance as metric of evaluation

The OCR-Free Document Understanding Transformer

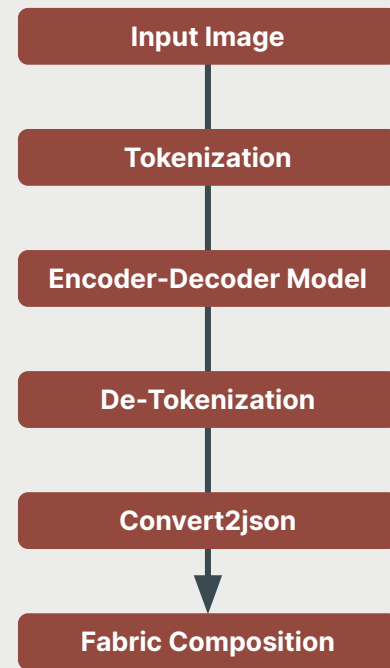


Encoder

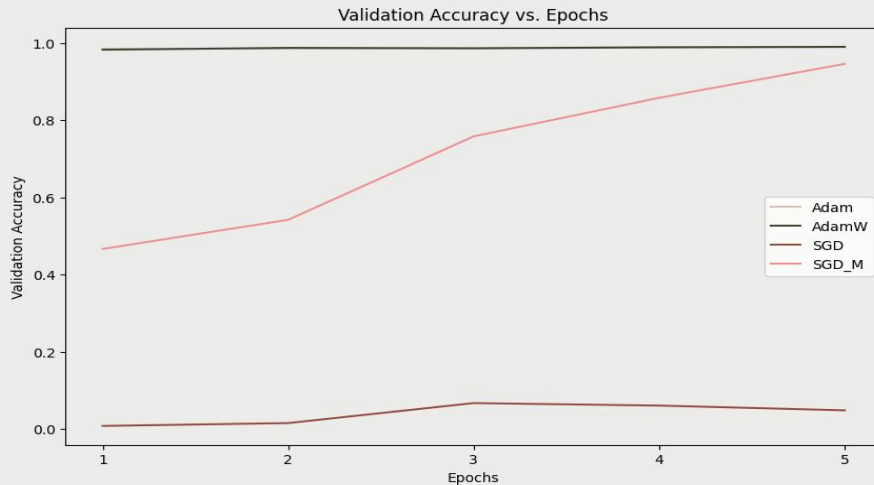
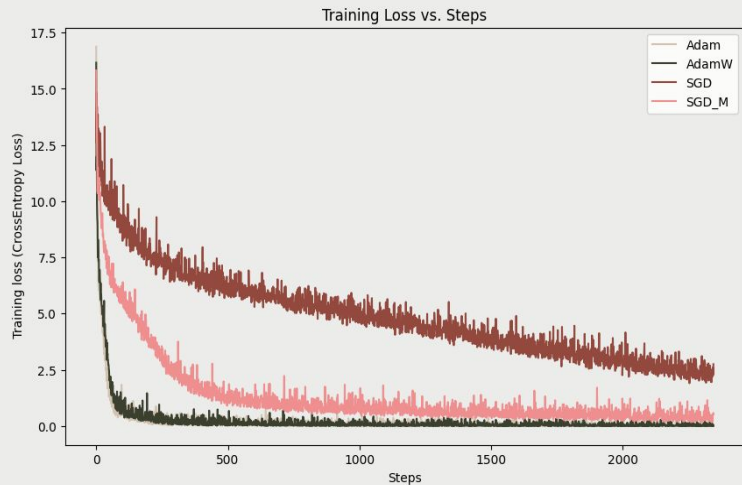
Prediction:
{
 'sample_id': '6534',
 'composition': {
 'polyester': '33.0', 'cotton': '67.0'}}



Decoder



Model Performance – *Hyperparameter Tuning*



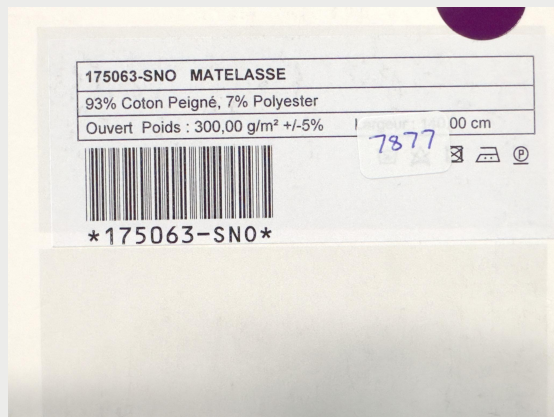
Optimizer	Samples with perfect match (Total Samples: 66)	Average Normalized Levenshtein Score (0-1)
Stochastic Gradient Descent (SGD)	0	0.95
SGD with Momentum	13	0.03
Adam	40	0.007
AdamW	47	0.005

Demo !



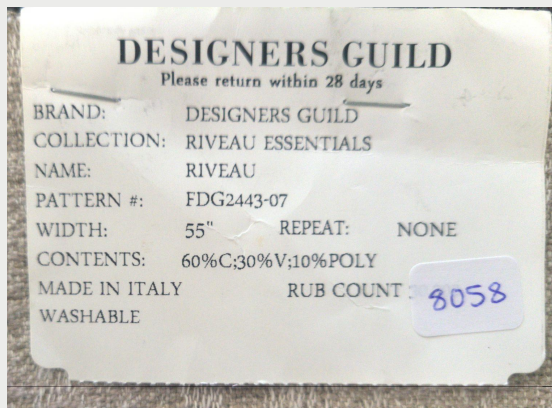
Model Robustness

Test Prediction 1



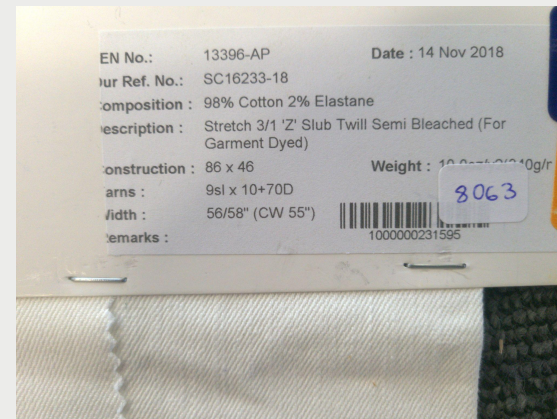
Prediction:
{ 'sample_id':
 '7877.0',
 'composition': {
 'polyester': '7.0',
 'cotton': '97.0' }}

Test Prediction 2



Prediction:
{ 'sample_id':
 '8058.0',
 'composition': {
 'polyester': '10.0',
 'cotton': '60.0',
 'viscose': '30.0' }}

Test Prediction 3



Prediction:
{ 'sample_id':
 '8063.0',
 'composition': {
 'cotton': '98.0',
 'elastane': '2.0' }}

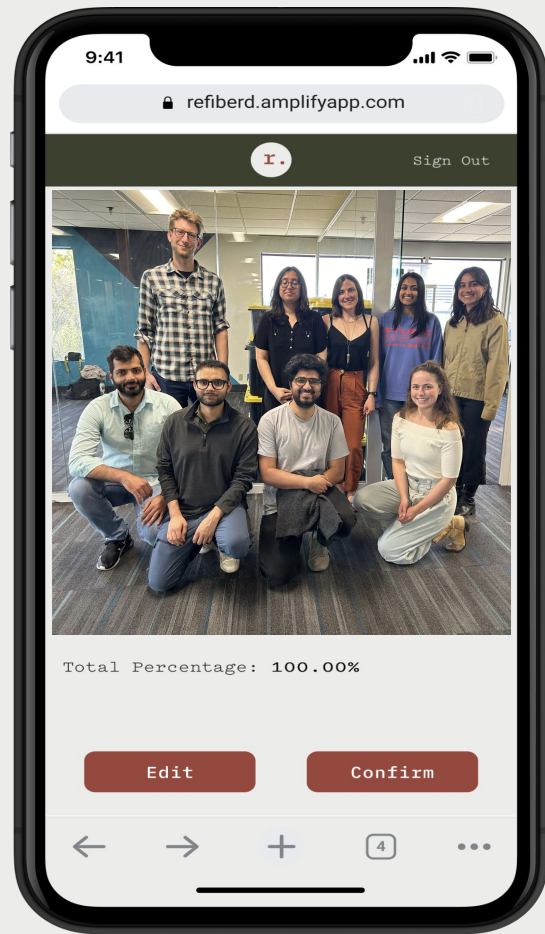
Key Learnings

We learned tremendously and delivered tangible results

Machine Learning will help in the fight against climate change by **expediting our fight**.

This project is a key step for helping in streamlined Refiberd's training data collection process, **a model is only as good as its training data**.

Learning **how to use AWS is critical** to set up a sound and scalable architecture the fully leverages its functionalities.





Isidora Rollan



Erin Jones



Mustafa Hameed



Prashant Sharma



Abdullah Azhar

Thank you

Refiberd Tag Reader
May 9th, 2024

refiberd.