

Quality Platform

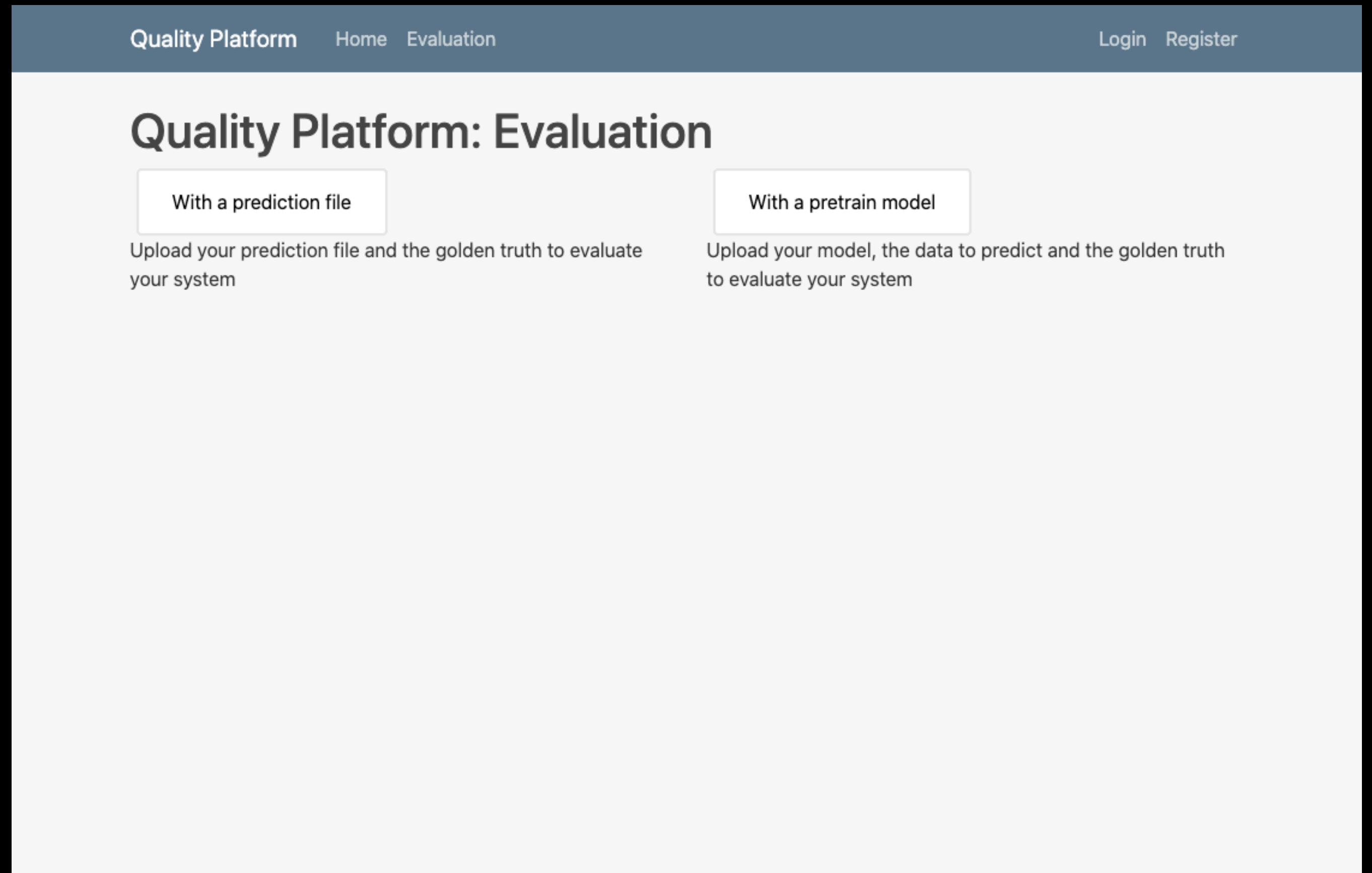
Gary Liu & Huazhang Liu

3 Functions

- Evaluate: evaluate user's language models, compare them and analyze the error
- Stress: use What-if scenarios to stress user's models and find improvements points
- Retrain: retrain user's models with the acquired knowledge and assign the proper weights

Evaluation

- With a prediction file
- With a pre-trained model (With Coco)



Evaluation Reports

- Overview
- Confusion Matrix
- Threshold Analysis
- Error Analysis
- Upload Another Predicted File (for evaluating with a predicted file mode only)

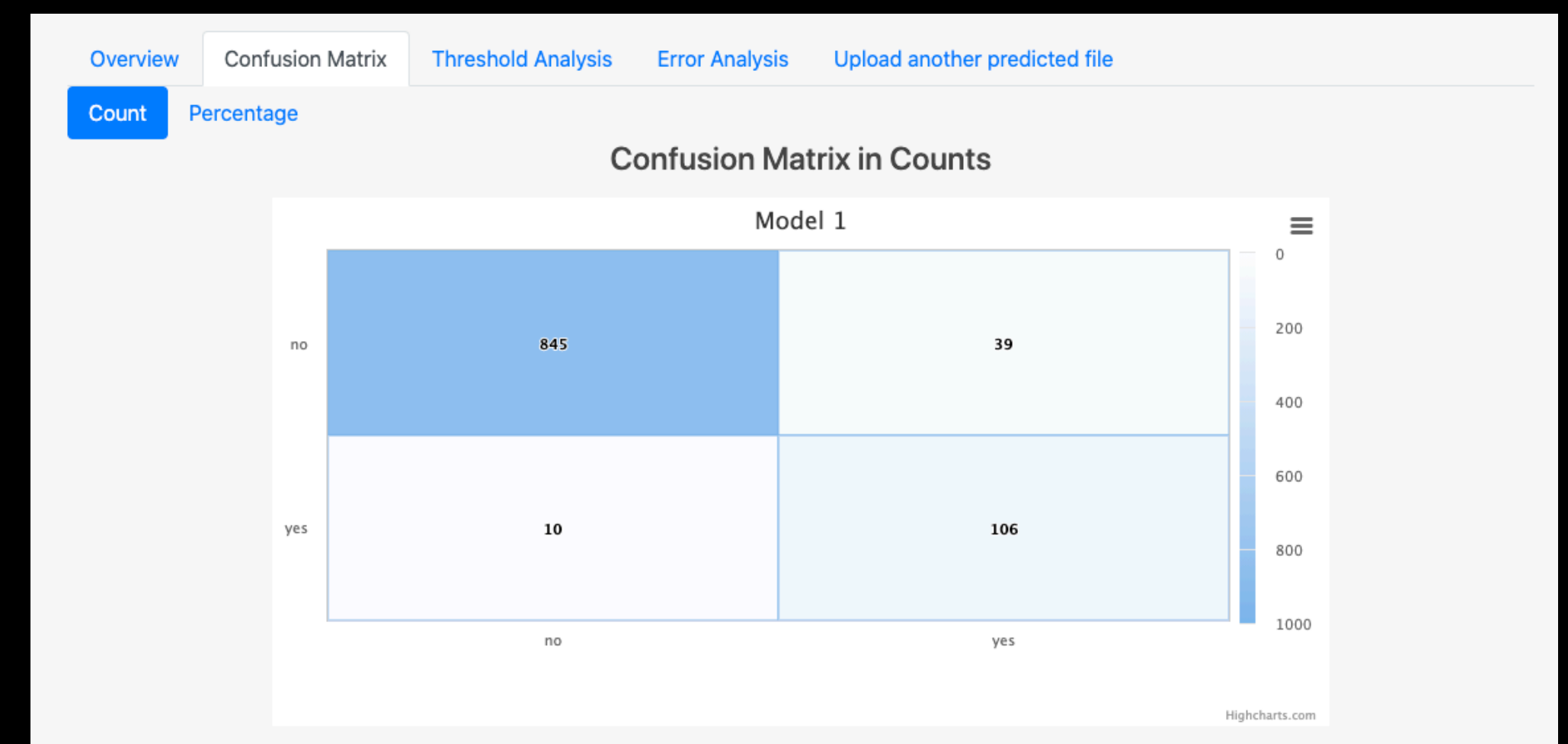
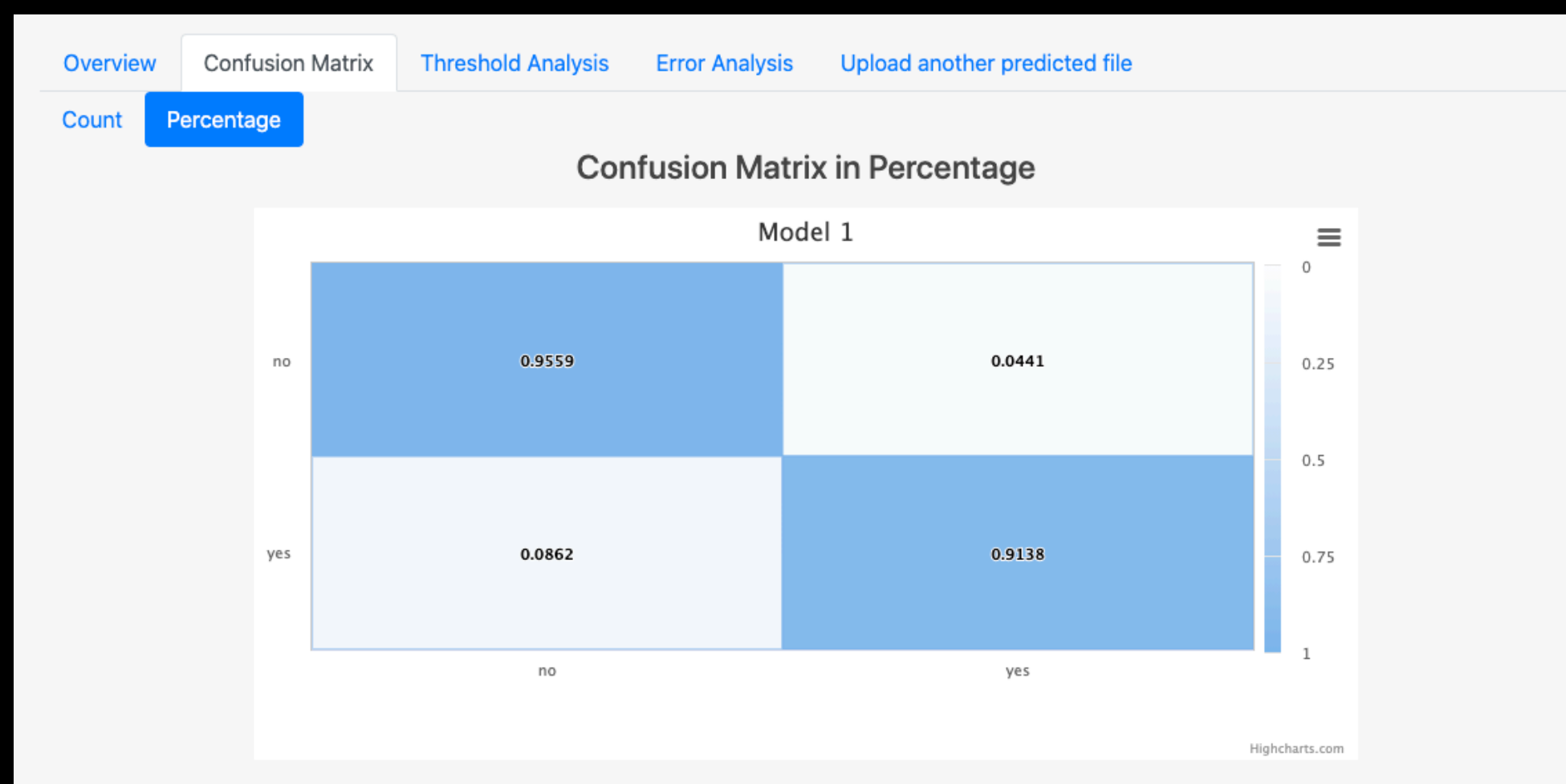
Overview

- An evaluation table containing accuracy, precision, recall, F1-scores, etc.

Quality Platform: Evaluation Report with a Prediction File				
With a prediction file		With a pretrain model		
Upload your prediction file and the golden truth to evaluate your system		Upload your model, the data to predict and the golden truth to evaluate your system		
Overview	Confusion Matrix	Threshold Analysis	Error Analysis	Upload another predicted file
	Precision	Recall	F1-score	Support
no	0.988	0.956	0.972	884
yes	0.731	0.914	0.812	116
accuracy			0.951	1000
macro avg	0.86	0.935	0.892	1000
weighted avg	0.958	0.951	0.953	1000
Total instance is 1000, and for each class:				
Class: no, Count: 884, Percentage: 0.884				
Class: yes, Count: 116, Percentage: 0.116				

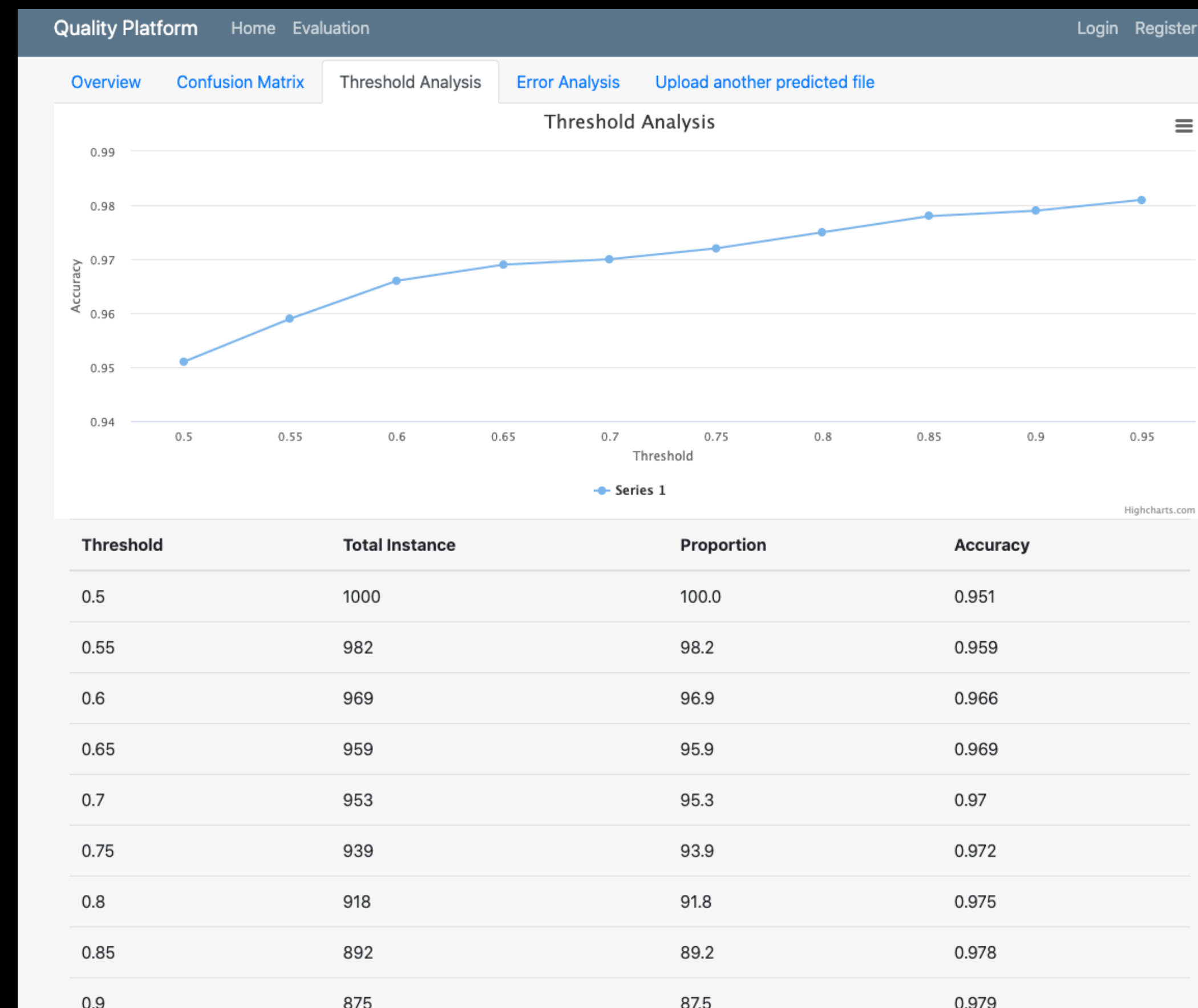
Confusion Matrix

- Heat maps of the predicted labels and the truth labels in count or percentage



Threshold Analysis

- Analyze the change of accuracy by increasing the threshold in 5% every time



Error Analysis

- A table containing wrong predicted data and its content

OverviewConfusion MatrixThreshold AnalysisError AnalysisUpload another predicted file

Search

<input type="checkbox"/>	ID	Actual Class	Predicted Class	Probability	Content
	<input type="text"/>	<div>no</div>	<div>yes</div>		<input type="text"/>
<input type="checkbox"/>	437d7d1d97917cd627a34a6a0fb41136	no	yes	0.99908	@karinpaynter @UberUK Imagine being sad enough to log on to Twitter and @ uber and complain . Ain't you got nothing better to do ?
<input type="checkbox"/>	f64eac11f2cd8f0efa196f8ad173178e	no	yes	0.99904	My phone call to the Macmillan Nurse did the trick ! Go in as a day patient on 25th for biopsy . Think you were right in saying shout or no one will hear you . On that day my friend Linda goes in for a Left Mastectomy as it did not respond to Chemo so we will both be in on the same day . We had a ""Cancer Hug"" this afternoon . They have given her a really good Wig , and unless you know you could not tell it was not hers !
<input type="checkbox"/>	39461a19e9eddfb385ea76b26521ea48	no	yes	0.99895	Hey josh , why do hypervenoms always rip at the front , please answer

Further expansion on this project

- Move the server running on local machine to run on cloud.
- Stressing the model with different test set.
- Retrain a new model after modification on weights and parameters.
- User group management and Admin management. This can be done in Django.

Thank you

And THANK YOU Kico and Yassine for guiding and helping us throughout this summer!