

Legal Intake & Machine Learning

Najmeh Forouzandehmehr,
Brendan Herger, Derek Walsh

Slide deck: <https://goo.gl/CVKkeM>

Intro
Project Features
Q+A

Intro

Currently, Capital One receives tens of thousands of subpoenas every year, and responds to them in a labor intensive and heuristic decision based process.

Intro

- Client: Legal Intake team
- Previous process: Highly manual & heuristic document sorting, triaging and information extraction
- New process: Automated, normalized document sorting, triaging and information extraction

Product Features

Feature	Outputs	Output frequency	Confidence level	Comment
Risk Bucketing	High Risk , Moderate Risk	One output per document	Probability from Naïve Bayes classifier	
Entity Served	e.g.: Capital One Bank (USA), N.A.	Zero or more outputs per document	None	regex match, based on normalized text
Document Type	Grand Jury , Credit Verification , Witness , Summons	One output per document	Probability from Naïve Bayes classifier	
Customer Identifiers	Social Security Number, EIN, Account Number	Zero or more outputs per document	Minimum OCR character confidence	

Product Features

Naive Bayes

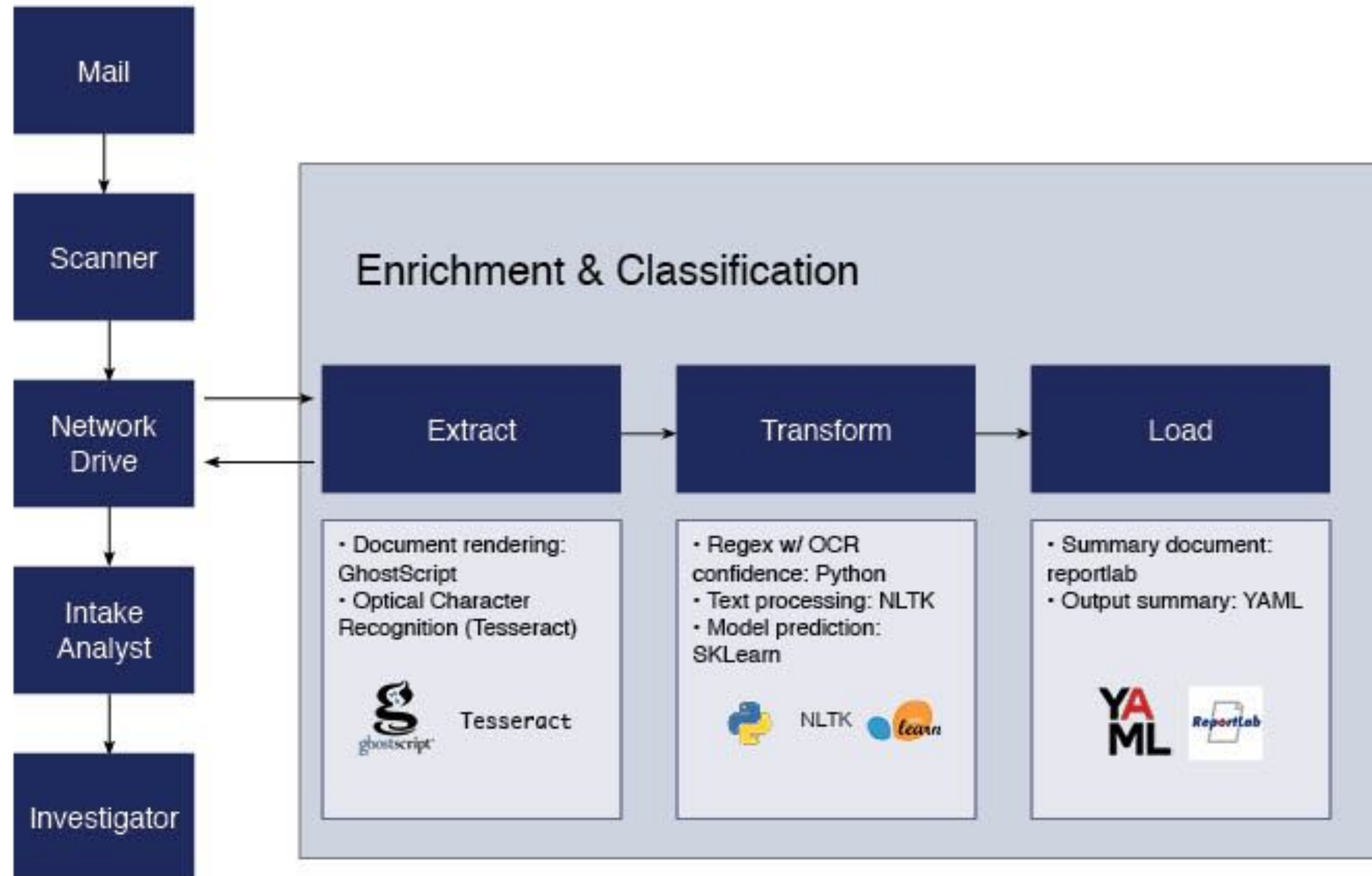
- Text processing
- Hyper-parameter selection (Grid Search + CV)
- Metrics evaluation

Product Features

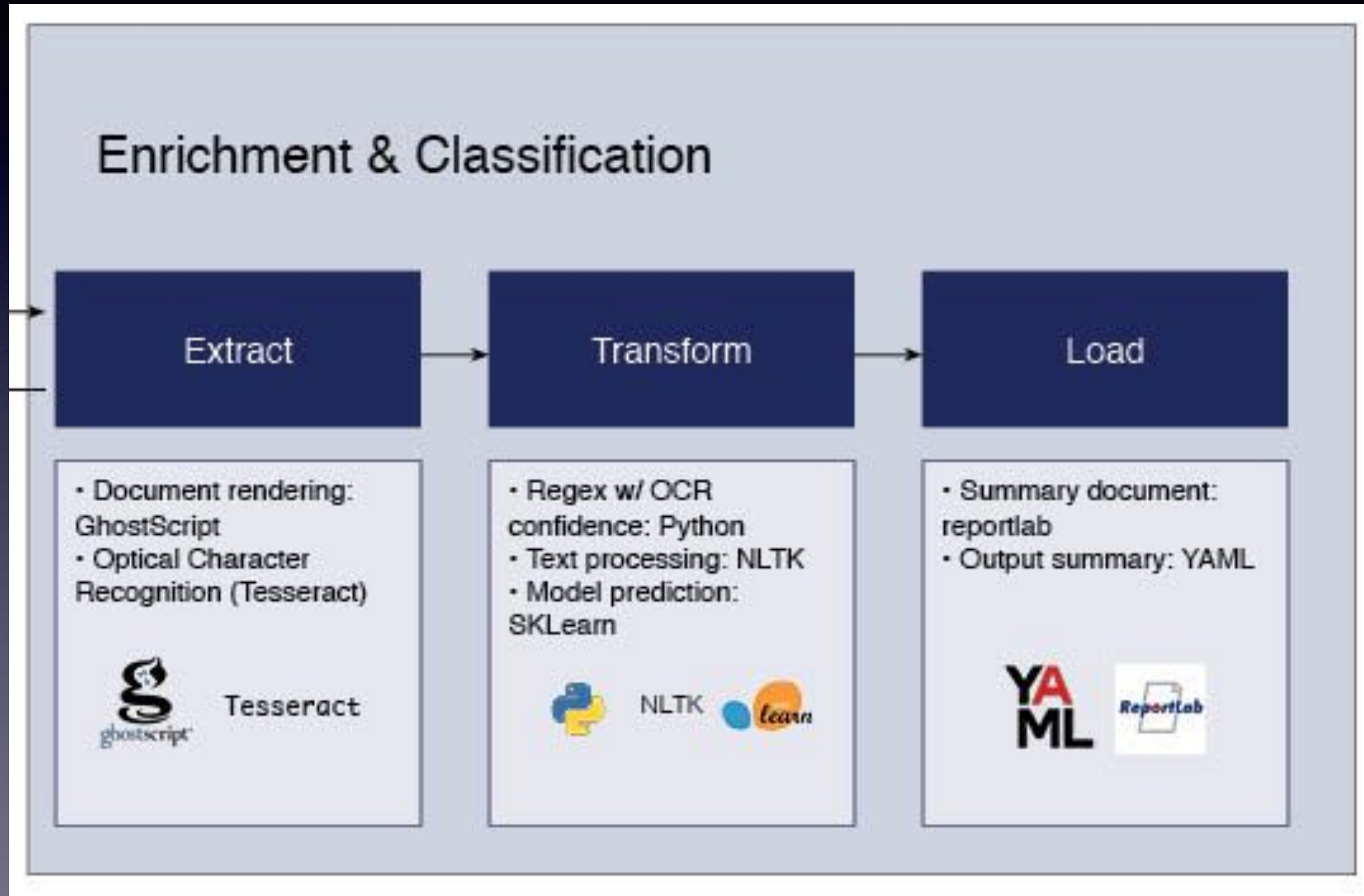
Regex w/ OCR Confidence

- OCR w/ per word confidence
- Format into table w/ index, character, confidence
- Find indices of matches
 - Match text is the text between these indices
 - Match confidence is an aggregate of the confidences between these indices

Product Features



Product Features



Q+A

Sound like fun? Come join us!

Me: Brendan.Herger@capitalone.com

Slide deck: <https://goo.gl/CVKkeM>