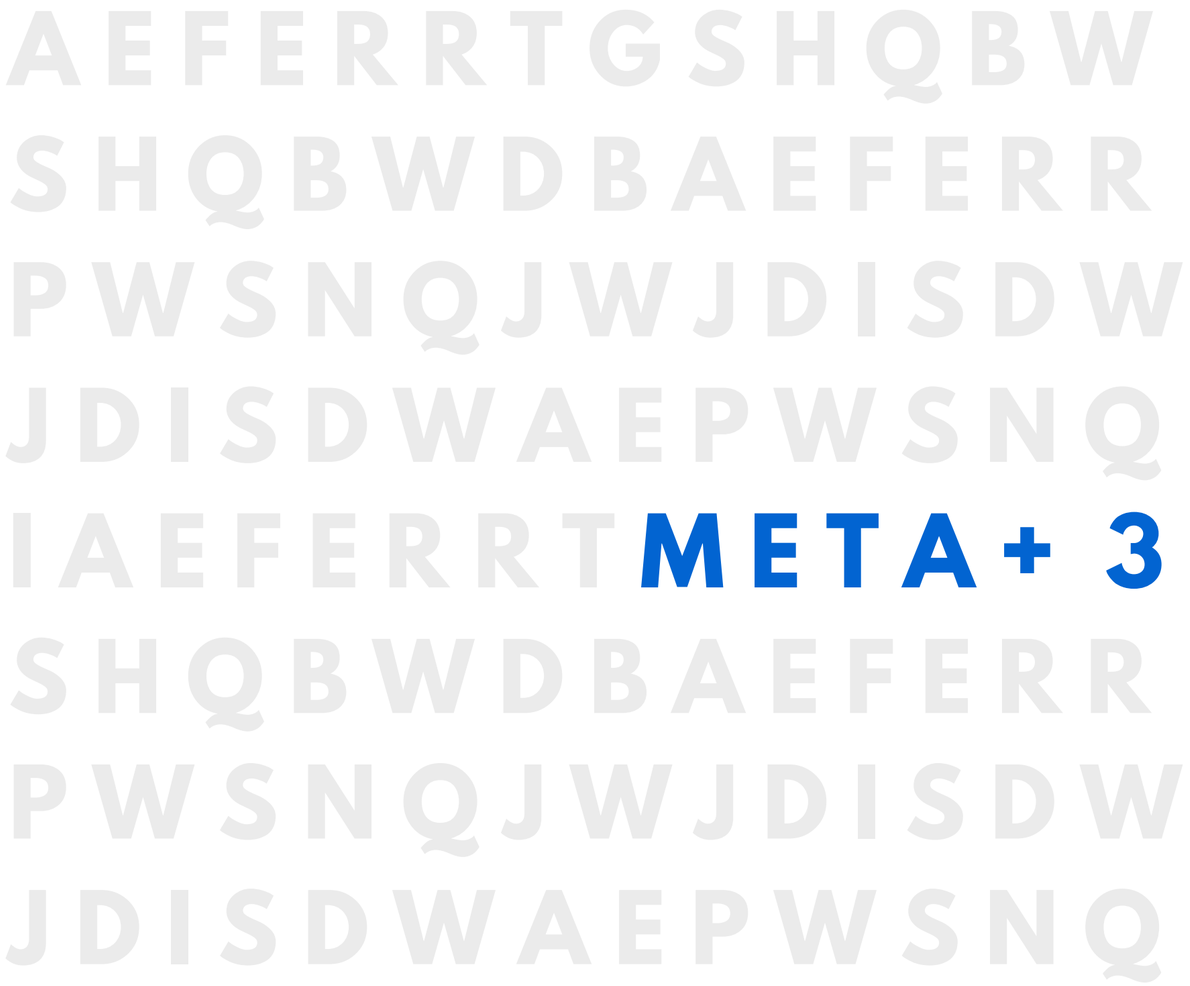


META + 3

SPRINT 4 - AI

Class: 3

Group: Meta+3

Project: Meta

CONTENT

1. Recap
2. Updates
3. Classification analysis
4. Error analysis
5. Our MVP
6. Next steps

RECAP

UPDATES

CONTENT

RECAP

UPDATES

CLASSIFICATION ANALYSIS

Recap

1
Project understanding
2
Architecture adaptations
3
Simulations and tests
4
Artificial Intelligence Classification
5
TBD

CONTENT

RECAP

UPDATES

CLASSIFICATION ANALYSIS

ERROR ANALYSIS

Updates

- Massive texts in English X Meta Database
- Lack of Target
- Supervised and unsupervised model
- Context API
- Use of Latam data
- Second Level Classification

RECAP

UPDATES

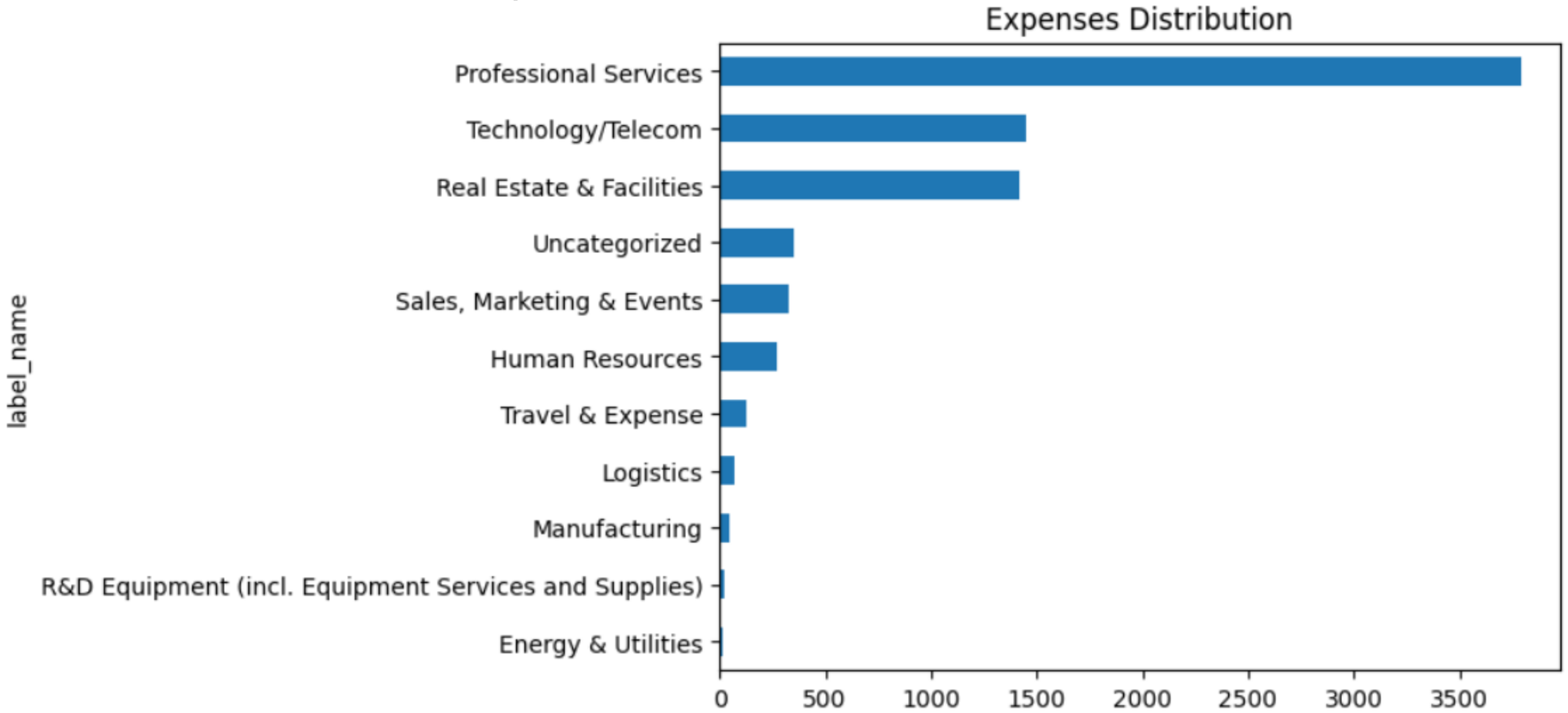
CLASSIFICATION ANALYSIS

ERROR ANALYSIS

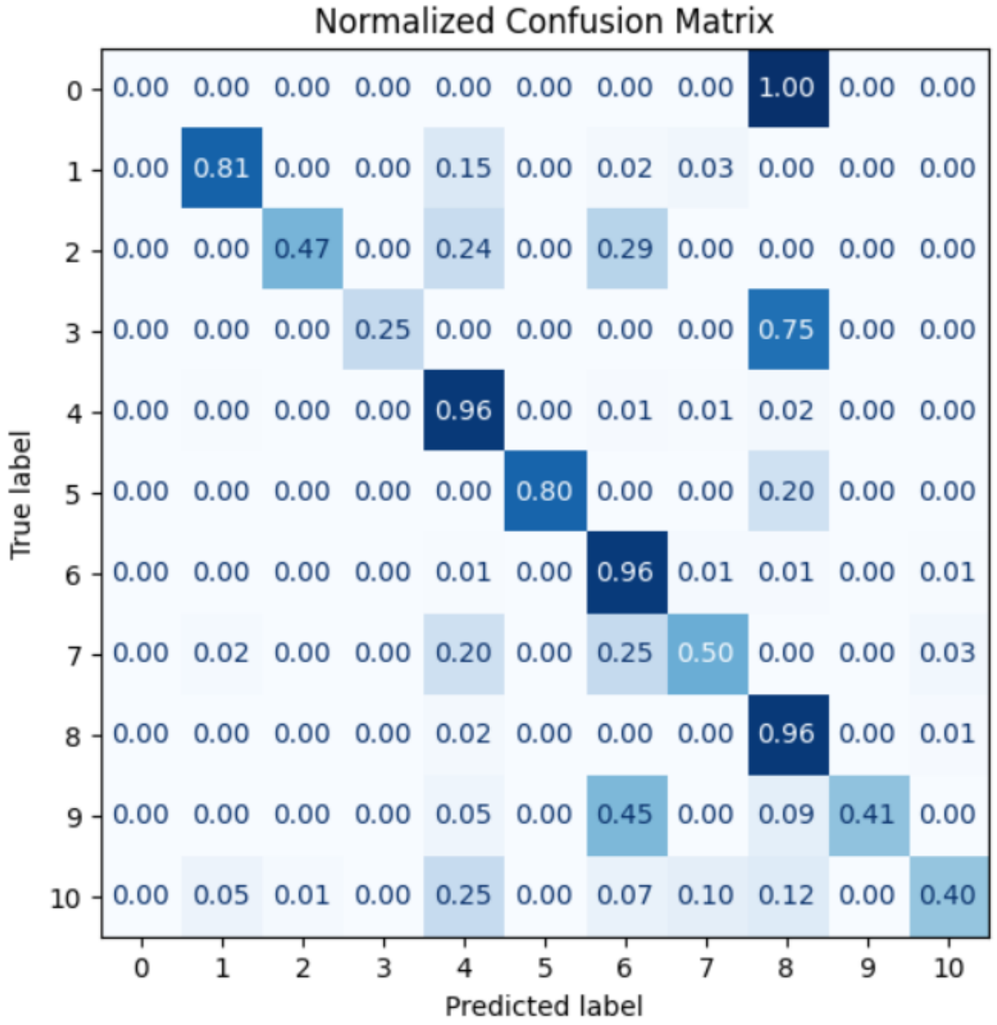
OUR MVP

Expenses Distribution

In this case, the class distribution is skewed towards the Professional Services class, which is the most frequent one.



Confusion Matrix




```
{'Energy & Utilities': 0,
 'Human Resources': 1,
 'Logistics': 2,
 'Manufacturing': 3,
 'Professional Services': 4,
 'R&D Equipment': 5,
 'Real Estate & Facilities': 6,
 'Sales, Marketing & Events': 7,
 'Technology/Telecom': 8,
 'Travel & Expense': 9,
 'Uncategorized': 10}
```

UPDATES

CLASSIFICATION ANALYSIS

ERROR ANALYSIS

OUR MVP

NEXT STEPS

Error Analysis

	text	label	predicted_label	loss
893	Normalized Supplier Name, ECOLAB, Country Name, Brazil, Strategic Region, NORAM, Requestor Name, Lisandro Ulerio, Preparer Name, Victória Simonato, Business Unit, Facilities, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), Non-Bay Area Facilities & Operations Group, Cost Center (Level 4), FCS G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Kitchen, Invoice Source, CAFEPRO, GL Description, Kitchen, Project, 31311 - Sao Paolo Aguas de Marco	Human Resources	6	5.146794
581	Normalized Supplier Name, WEW IMPORTACAO E EXPORTACAO LTDA, Country Name, Brazil, Strategic Region, NORAM, Requestor Name, Lisandro Ulerio, Preparer Name, Victória Simonato, Business Unit, Facilities, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), Non-Bay Area Facilities & Operations Group, Cost Center (Level 4), FCS G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Kitchen, Invoice Source, CAFEPRO, GL Description, Kitchen, Project, 31311 - Sao Paolo Aguas de Marco	Logistics	6	5.145047
354	Normalized Supplier Name, BRIGAGAO DUQUE ESTRADA ADVOGADOS, Country Name, Brazil, Strategic Region, LATAM, Requestor Name, Rafael Benevides, Preparer Name, Rafael Benevides, Business Unit, Finance, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), Tax Group, Cost Center (Level 4), Finance G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Audit and tax services, Invoice Source, LETTERBOX, GL Description, Audit and tax services, Project, 00000 - Default Project	Uncategorized	4	4.738184
1322	Normalized Supplier Name, LENOVO TECNOLOGIA BRASIL LIMITADA, Country Name, Brazil, Strategic Region, NORAM, Requestor Name, Stephen Murphy, Preparer Name, ERIKA BOVO, Business Unit, IT & Corpnet, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), EE Ops Supply Chain Group, Cost Center (Level 4), Enterprise Engineering G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Fixed Assets - Gross, Invoice Source, LETTERBOX, GL Description, Assets clearing-capitalized, Project, 00000 - Default Project	R&D Equipment (incl. Equipment Services and Supplies)	8	4.605190
360	Normalized Supplier Name, DIOGO RAIS RODRIGUES MOREIRA, Country Name, Brazil, Strategic Region, LATAM, Requestor Name, unknown, Preparer Name, unknown, Business Unit, Legal, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), NORAM Litigation Group, Cost Center (Level 4), Legal G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Legal - litigation, Invoice Source, LETTERBOX, GL Description, Legal - litigation, Project, 00000 - Default Project	Uncategorized	4	4.482581

- The goal of this exercise is to detect wrong labels or quirks of the dataset.
- With this information we can refine the dataset, which might be either mislabeled or require a new class altogether.
- Deep learning models are exceptionally good at finding and exploiting shortcuts to get to a prediction.
- Data samples with highest losses

CLASSIFICATION ANALYSIS

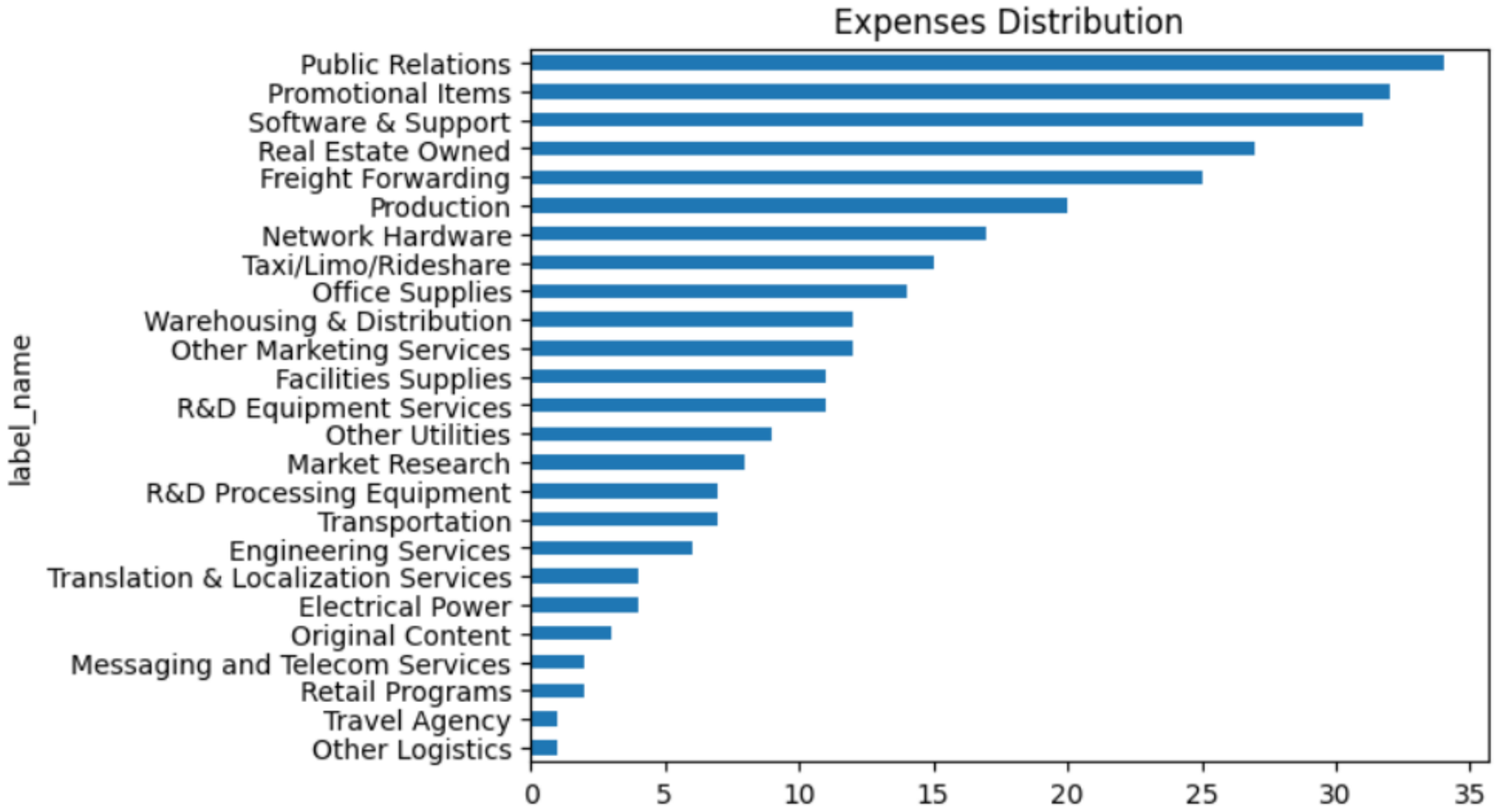
ERROR ANALYSIS

SECOND LEVEL CLASSIFICATION

NEXT STEPS

Expenses Distribution

In this case, the class distribution is skewed towards the Professional Services class, which is the most frequent one.



CLASSIFICATION ANALYSIS

ERROR ANALYSIS

CONCEPT APLICATION

NEXT STEPS

Second Level Classification

- We refined the dataset that may be poorly labeled.
- Consequently, we discovered shortcuts to achieve better predictions precisely because we investigated which predictions the model is more confident in."

```
'13': 'Expedited Freight',
'14': 'Facilities Services',
'15': 'Facilities Supplies',
'16': 'Financial Services',
'17': 'Food & Beverage',
'18': 'Freight Forwarding',
'19': 'Full Service Creative Agency',
'20': 'Insurance',
'21': 'Legal Services',
'22': 'Market Research',
'23': 'Media',
```

	text	label	predicted_label	loss
498	Normalized Supplier Name, PIRAMIDES COMERCIO DE AGUAS MINERAIS E BEBIDAS EM GERAL LTDA, Country Name, Brazil, Strategic Region, NORAM, Requestor Name, Lisandro Ulerio, Preparer Name, Victória Simonato, Level 1, Real Estate & Facilities, Business Unit, Facilities, Legal Entity, Facebook Servicos Online Do Brasil Ltda, Cost Center (Base Level), Non-Bay Area Facilities & Operations Group, Cost Center (Level 4), FCS G&A, Cost Center (Level 5), General & Administrative, GL Desc (Level 4), Kitchen, Invoice Source, CAFEPRO, GL Description, Kitchen, Project, 31311 - Sao Paolo Aguas de Marco	Real Estate Owned	17	5.629463

CLASSIFICATION ANALYSIS

ERROR ANALYSIS

OUR MVP

NEXT STEPS

ERROR ANALYSIS

OUR MVP

NEXT STEPS

Next Steps

1
2
3
4
5
Project understanding
Architecture adaptations
Simulations and tests
Artificial Intelligence Classification
TBD

THANK YOU!