



Machine Learning Utilisation in Call Categorisation

Task

To apply a machine learning algorithm to call transcripts to categorise calls.

Working with a labelled training set, to train a classification algorithm to then work with unlabelled data.

Category 6 Calls

- Introducing the customer to the company and process
- Gathering customer preferences
- Briefing customers on loan application process

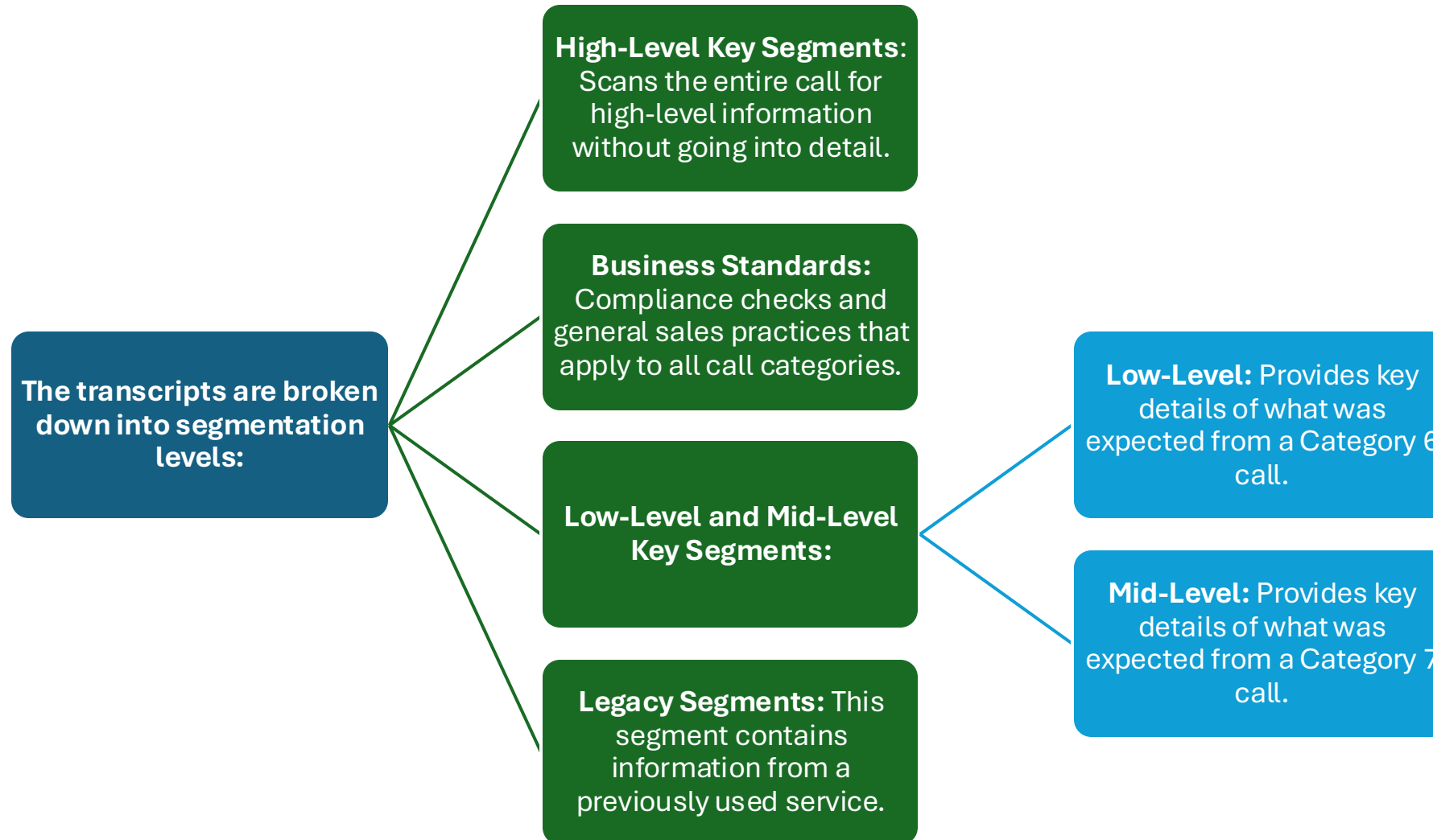
Category 7 Calls

- Finalise loan application
- Explaining further customer benefits
- Clarifying next steps



Overview of Data

The call transcripts are stored in a large language model format



Overview of Data

Call_Type_Name	Low_Level_Segment6
Category 1	{'item1': 'Red', 'item2': 'Red', 'item3': 'Red', 'item4': 'Red', 'item5': 'Red', 'item6': 'uncategorized'}
Category 4	{'item1': 'Red', 'item2': 'Red', 'item3': 'Red', 'item4': 'Red', 'item5': 'Red', 'item6': 'uncategorized'}
Category 7	{'item1': 'Red', 'item2': 'Red', 'item3': 'Red', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Red', 'item2': 'Green', 'item3': 'Red', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Red', 'item2': 'Red', 'item3': 'Green', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Red', 'item2': 'Green', 'item3': 'Green', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Green', 'item2': 'Red', 'item3': 'Red', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Red', 'item2': 'Red', 'item3': 'Red', 'item4': 'Green', 'item5': 'Red', 'item6': 'initial_call'}
Category 7	{'item1': 'Red', 'item2': 'Green', 'item3': 'Green', 'item4': 'Green', 'item5': 'Green', 'item6': 'initial_call'}
Category 6	{'item1': 'Green', 'item2': 'Green', 'item3': 'Green', 'item4': 'Green', 'item5': 'Green', 'item6': 'initial_call'}

The columns contain dictionaries instead of simple values, making it difficult to process

Machine learning models expect structured tabular data, not nested dictionaries

The dataset includes text labels like "Red", "Green", "uncategorized", which are not machine-readable

These need to be converted into numerical features for machine learning

Our Solution

Call_Type_Name	Low_Level_Segment6_item1	Low_Level_Segment6_item2	Low_Level_Segment6_item3	Low_Level_Segment6_item4	Low_Level_Segment6_item5	Low_Level_Segment6_item6
Other	0	0	0	0	0	3
Other	0	0	0	0	0	3
Category 7	0	0	0	2	0	0
Category 7	0	2	0	2	0	0
Category 7	0	0	2	2	0	0
Category 7	0	2	2	2	0	0
Category 7	2	0	0	2	0	0
Category 7	0	0	0	2	0	0
Category 7	0	2	2	2	2	0
Category 6	2	2	2	2	2	0

Flattened nested data by converting dictionary values into separate columns

Ensured consistent formatting so all data follows a structured tabular format

Converted categorical text to numbers using Label Encoding

Converted Category 1-5 calls to ‘Other’



Supervised Model

A fully supervised classification algorithm that can be trained with labelled data and categorise call types between category 6 and 7 calls

Creating a **Random Forest** algorithm to provide classification

Utilising its decision tree architecture to involve all segmentation levels in its pursuit of categorising calls

Each segmentation level populating a different decision tree within the classification algorithm

Incorporating all the data in final prediction and operates well when working with large complex datasets

Category Predictions

Category	Count
Other	2121
Category 7	714
Category 6	56

Evaluation of model:

Category	Precision	Recall	F1-Score	Support
Category 6	1.0	0.92	0.96	12
Category 7	0.89	1.0	0.94	56
Other	1.0	0.5	0.67	12
Accuracy			0.91	80
Macro Avg	0.96	0.81	0.85	80
Weighted Avg	0.92	0.91	0.9	80



Semi-Supervised Model

Augmenting the designed random forest algorithm to become semi-supervised

This change allows for the algorithm to then label its own data and not requiring the training with such a large labeled dataset

Category Predictions

Category	Count
Other	2159
Category 7	675
Category 6	57

Evaluation of model:

Category	Precision	Recall	F1-Score	Support
Category 6	1.0	0.92	0.96	12
Category 7	0.88	1.0	0.93	56
Other	1.0	0.42	0.59	12
Accuracy			0.9	80
Macro Avg	0.96	0.78	0.83	80
Weighted Avg	0.91	0.9	0.89	80