Customer Feedback Monitoring

📋 Overview

We work closely with CX/CS teams to employ our in-house NLP based solutions for better monitoring of written-text customer feedbacks about Getir and its domains. This documentation page aims to provide reference and unify all available resources of past projects.

How does this project fit into your broader strategy?	Improve written-text feedback monitoring for automated customer services.
Team	DS4CX Team
Members	@Çağla Özen @Hasan Can Karapınar @Emre Selçuk @Dorukhan Afacan

❓ Tackled Problems

For detailed look click on project titles and dataset links.

Title	Status	Source Table	Output Table
Market Order Comments Classification (UK included)	DONE	marketorders, etl_market_rating.marketratings	etl.comment_analytics.comment_analysis
US Comment Classification	DONE	marketorders	us_comment_classification
EkşiSözlük Entry Classification	BLOCKED	eksisozluk_crawling	eksisozluk_comment_classification
ŞikayetVar Complaint Classification	BLOCKED	sikayetvar_crawling	sikayetvar_comment_classification
Twitter Tweet Classification	DONE	Twitter API	twitter_comments_classification
Global InApp Reviews	DONE	etl_market_rating.marketratings	multilang_comment_classification
Artisan Comment Classification	DONE	etl_artisan_order.foodorders	artisan_comment_classification
BiTaksi Comment Classification	DONE	etl_getir_bitaksi.rates	bitaksi_comment_analytics
Moov Comment Classification	DONE	etl_moov_skybackend.rent_rentfeedback	moov_comment_analysis
Getir Water Marketplace Comment Classification	DONE	TBD	TBD
GetirJobs Content Filtering	DONE	etl_getirjobs_backend	--
Sensor Tower	DONE	Sensor Tower API	app_comments_reviews
PreOrder Notes	DONE	TBD	TBD
Locals Courier Comment Classification	DONE	TBD	locals_courier_comment_classification
UK NPS Survey	DONE	Google Sheets	TBD
Global NPS Survey	DONE		
Post Agent CSAT Surveys	IN PROGRESS		
Store Reviews Topics	IN PROGRESS		

🤦‍♂ Tools and Stack

An NLP MLOps cycle in currently under development to carry existing and incoming textual feedback classification tasks. The aim is to downsize all NLP problems into **data curation** problems s.t. solutions will be data centric instead of being model/code centric. This will help two major issues:

- eliminate the boiler-plate development each time there is a new task. - Data Side
- offer monitoring and explainability for stakeholders to collect instances for re-tranining. - CX Side.
- provide a centralized standard for implementation, reproducibility and avoid information loss for future teams.

The key elements in the cycle are:

Data Annotation

Manual Annotation

LLM Assisted Auto Annotation

Inference Time

There are many repeating elements in inference Cron jobs of past NLP models. An inference template cron is prepared for east deployment of scheduled models.

Model Training and Optimization

Current Codebase

Targeted Codebase

Less model code. More data curation.

Experiment Tracking

Experiment tracking screen from W&B

Explainability - FUTURE

Project Details

Market Order Comment Classification*

- Multi-label classification model for tagging reviews at `rate_comment` field of `marketorders` table. A language detection model dispatches Turkish reviews to this model. Reviews with 1,2,3 star rating are passed through the model. A coarse level categorization is applied to instances with a binary classification model (`other`, `non-other`). `non-other` instances flow through multi-label classification model. For more info and categories refer to above data sources.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	
Cron Link	Cron link.
Model Link	Binary model link, Model Link.
Tableau Link	Tableau Link.

US Comment Classification

- Multi-label classification model for tagging reviews at `rate_comment` field of `marketorders` that originate from United States. Reviews with 1,2,3 star rating are passed through the model. For details of class cardinality refer to above data sources. For calculation of the multi-label metrics refer to this link.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Instance-Precision:** 0.8950 **Instance-Recall:** 0.9018 **Instance-F1:** 0.8920 **Macro-F1:** 0.8767
Cron Link	Cron link.

Model Link	Model Link.

UK Comment Classification

- Multi-label classification model for tagging reviews at `rate_comment` field that originate from United Kingdom in `marketorders` table. Reviews with 1,2,3 star rating are passed through the model. A coarse level categorization is applied to instances with a binary classification model (`other` , `non-other`). `non-other` instances flow through multi-label classification model. For more info and categories refer to above data sources.

Codebase	Repo Link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	--
Cron Link	--
Model Link	Binary Model link, Model link.
Tableau Link	--

EkşiSözlük Entry Classification

- Entries from various **EkşiSözlük** titles are classified into 14 categories. Titles are related to Getir, its domains, commercials and competitors. Shares the same model with **Şikayet Var.** For details and class cardinality refer to above data sources.

Codebase	Repo link.
Scraper	Crawler repo link.
Sentiment Included	FALSE
Eval Metric(s)	**F1:** 0.9465 **Precision:** 0.9463 **Recall:** 0.9470
Cron Link	Cron link.
Model Link	Model Link.
Tableau Link	Tableau Link.

Şikayet Var Complaint Classification

- Complaints from **Şikayet Var** are classified into 14 categories. For details of class cardinality refer to above data sources. Shares the same model with **EkşiSözlük** Currently in BLOCKED status due to API ban.

Codebase	Repo link.
Scraper	Crawler repo link.
Sentiment Included	FALSE
Eval Metric(s)	**F1:** 0.9465 **Precision:** 0.9463 **Recall:** 0.9470
Cron Link	Cron link.
Model Link	Model Link.
Tableau Link	Tableau Link.

Twitter Tweet Classification

- Getir twitter mentions that are crawled via `twint` API are passed through downstream tasks of sentiment analysis and comment classification. Negative labeled instances are dispatched to classification model to be classified into 16 categories. For details and class cardinality refer to above data sources.

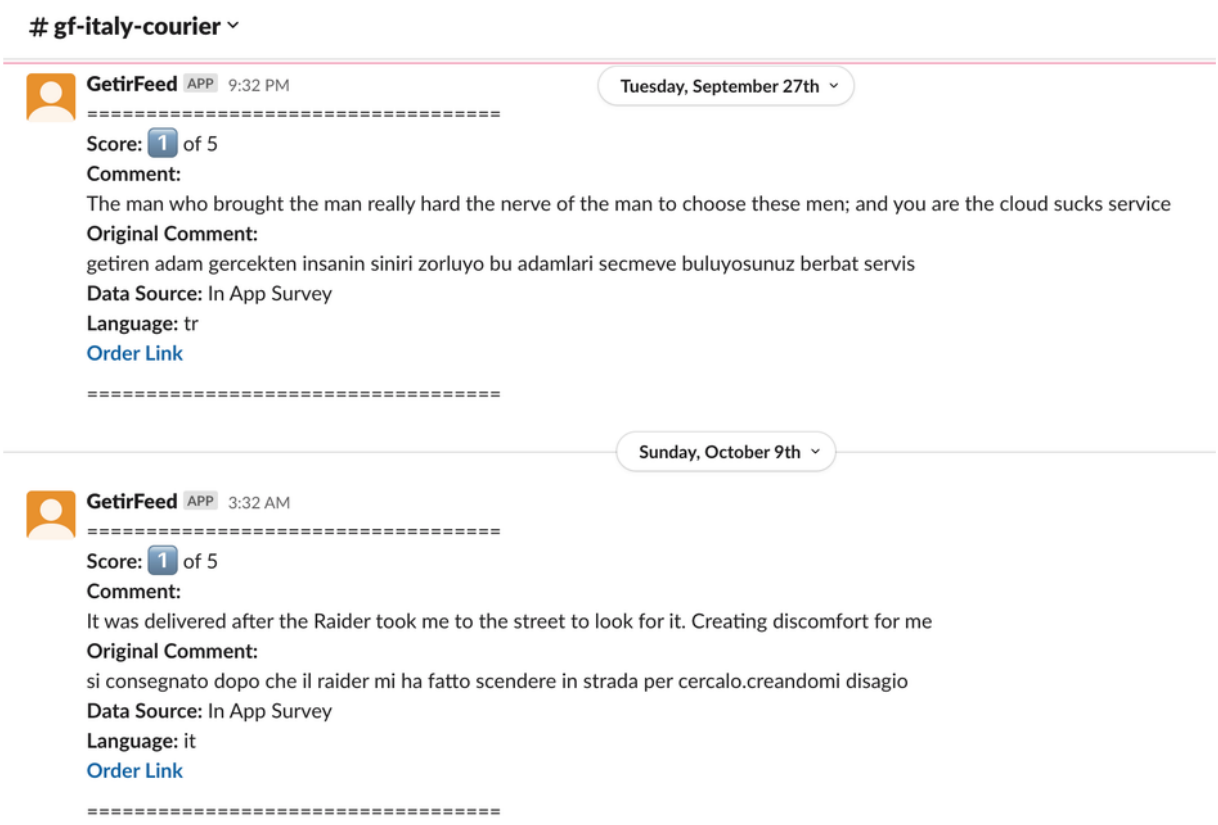
Codebase	Repo link.
Scraper	--
Sentiment Included	TRUE
Eval Metric	--
Cron Link	Cron link.
Model Link	Model Link. Sentiment Model link.
Tableau Link	Tableau Link.

Birdie (Chattermill In-house replacement)

- **Chattermill** is an outsource tool for customer feedback analytics. Data flowing through Chattermill consists of international reviews from 65+ languages. They are translated to English before going through a separate sentiment analyzer and categorization model. NLTK `SentimentInstensityAnalyzer` is used for sentiment classification. An in-house multi-label classification model with 11 categories is used for tagging reviews. For calculation of the multi-label metrics refer to this link.

Codebase	Repo link.
Scraper	--
Sentiment Included	TRUE
Eval Metric	**Instance-Precision:** 0.8988 **Instance-Recall:** 0.9059 **Instance-F1:** 0.9006 **Macro-F1:** 0.8724
Cron Link	Cron link.
Model Link	Model Link.
Tableau Link	Tableau Link.

- NonTR In-app reviews are labeled by classification model and sentiment analyzer. Later on they are passed through filters/workflows(e.g. Country: France, Label: expired product, sentiment: negative) created by local CX teams.



SlackBot posts Reviews that pass through workflows.

Artisan Comment Classification

- Classification model for tagging reviews at **GetirÇarşı with 12 in-house labels.** Reviews with 1,2,3 star ratings are passed through the model. A coarse level categorization is applied to instances with a binary classification model (`other` , `non-other`). `non-other` instances flow through multi-label classification model. For more info and categories refer to above data sources.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric	--
Cron Link	Cron link.
Model Link	Binary classification model, Classification model Models Link.
Tableau Link	Tableau Link.

BiTaksi Comment Classification

- **BiTaksi** trip reviews with 1,2,3 star ratings are tagged with a multi-label classification model with 10 categories. For details and class cardinality refer to above data sources. For calculation of the multi-label metrics refer to this link.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Instance-Precision:** 0.7243 **Instance-Recall:** 0.7947 **Instance-F1:** 0.7331

	Macro-F1: 0.5618
Cron Link	Cron link.
Model Link	Model Link.
Tableau Link	Tableau Link.

Moov Comment Classification

- Driving Experience reviews with 1,2,3 star rating are tagged with a multi-label classification model that has 13 categories. For calculation of the multi-label metrics refer to this link.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Instance-Precision:** 0.8738 **Instance-Recall:** 0.8903 **Instance-F1:** 0.8693 **Macro-F1:** 0.8147
Cron Link	Cron link.
Model Link	Model Link.
Tableau Link	--

GetirWaterMarketPlace Comment Classification

Codebase	Repo link.
Scraper	--
Sentiment Included	--
Eval Metric	**Accuracy:** 0.9640 **Precision:** 0.8389 **Recall:** 0.8430 **Macro-F1:** 0.8403
Cron Link	TBD
Model Link	Model Link.
Tableau Link	TBD

GetirJobs Content Filtering

- Binary classification model to identify offensive language in GetirJobs postings. Model is trained with HemenİŞ dataset, SemEval 2020 Task12 Turkish language dataset and augmentations.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric	**F1 (NOTOFF):** 0.98, **F1 (OFF):** 0.91 **Recall (NOTOFF):** 0.99, **Recall (OFF):** 0.86 **Precision (NOTOFF):** 0.98, **Precision (OFF):** 0.96 **Total Accuracy:** 0.98
Cron Link	Panel link.
Model Link	Model Link.
Tableau Link	Tableau Link.

Sensor Tower

Multi-label classification on Sensor Tower App data accessed through API. Only Turkish reviews that have 1,2,3 star rating are used for classification. For calculation of the multi-label metrics refer to this link. Currently in BLOCKED status due to API ban.

Codebase	Repo link.
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Instance-Precision:** 0.84 **Instance-Recall:** 0.81 **Instance-F1:** 0.81 **Macro-F1:** 0.77
Cron Link	--
Model Link	Model Link.
Tableau Link	Tableau Link.

Pre-Order Notes Classification

Multi-label classification on pre-order customer notes. There are **21 in-house classes**. Data is being labeled using our in-house annotation tool. It is the pilot project for the MLOps pipeline.

Codebase	--
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Samples-F1:** 0.85 **Macro-F1:** 0.61 **Weighted-F1:** 0.84
Cron Link	--
Model Link	--
Tableau Link	--

Locals (Artisan) Courier Comment Classification

Multi-label classification of post-order customer comments about couriers in GetirLocals domain. Data is labeled by outsourced human annotators with **26 in-house classes**. Dataset link.

Codebase	**TBA**
Scraper	--
Sentiment Included	FALSE
Eval Metric(s)	**Samples-F1:** 0.85 **Macro-F1:** 0.57 **Weighted-F1:** 0.82
Cron Link	TBA
Model Link	TBA
Tableau Link	--

UK NPS Survey

The meaning of **NPS** is "*Net Promoter Score*". Analysis of comments in the **UK** is being done. An email survey is sent to Getir customers, where they can give a score between 1-10, write their suggestions and complaints. Surveys are sent to customers who have used the service for 1, 3, 6, and 12 months, and the results are recorded in Google Sheets accordingly.

The descriptions of score ranges;

- 9-10 -> *Promoter*
- 7-8 -> *Passive*
- <6 -> *Detractor*

The aim of this project is to use machine learning to automatically tag comments, instead of manually checking each comment to determine which category it belongs to, as the answers to the questions are open-ended. This is to reduce time waste. Because there are customers who are promoters but make negative comments in the suggestion section, or detractors who make positive comments in the suggestion section.

Four models were used in the project, which are suggestion_category, suggestion_sentiment, reason_category, and reason_sent. The ***"distilbert-base-uncased"*** model was used for sentiment, and the ***"roberta-base"*** model was used for category. The predictions are written to Athena.

Codebase	Repo link.			
Scraper	--			
Sentiment Included	TRUE			
Eval Metric(s)	REASON SENTIMENT **Accuracy-F1:** 0.95 **Macro-F1:** 0.89 **Weighted-F1:** 0.95	REASON CATEGORY **Micro-F1:** 0.92 **Macro-F1:** 0.82 **Weighted-F1:** 0.91 **Samples-F1:** 0.91	SUGGESTION SENTIMENT **Accuracy-F1:** 0.92 **Macro-F1:** 0.91 **Weighted-F1:** 0.92	SUGGESTION CATEGORY **Micro-F1:** 0.91 **Macro-F1:** 0.66 **Weighted-F1:** 0.89 **Samples-F1:** 0.91
Cron Link	Cron link.			
Model Link	Model Link.			
Tableau Link	--			