Model Validation Report

[DOCUMENT SUBTITLE]

**Document Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
| Version | Description of change | Author | date |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Model Info**

|  |  |
| --- | --- |
| Model | Empirical Cumulative Distribution based Outlier Detection |
| Model Version | 1.0 |
| Date |  |
| Primary Validator | Navid Kalantari |
| Model Registry link |  |

**Scope**

The following encompasses the scope of the MVR:

* An independent assessment of the appropriateness & performance of a model to meet its design objectives
* Observations of technical model risks and/or limitations, with associated mitigation steps taken
* Business risks associated with the model usage and the mitigations to those risks
* Model validator sign off that required and specified procedures for developing model were followed
* Summarize results of executing assurance plan
* Summarize other mode validation results

مواردی که هایلایت شده رو قبلا جیمز نوشته بود برای ولیدیشن خودش

**Model Performance/Appropriateness**

**Summary:**

Empirical Cumulative Distribution based Outlier Detection model looks at each scaled input and estimates an empirical cumulative probability function and then creates a left/right\_end tail probability per each broker firm. Then, it combines all by making independence assumption to generate the combined outlier score.

There are two separate outputs from firm-centric view:

* The first is the monthly anomaly score and rank per peer-group. The users are interested in looking at the top firms from each peer group/cluster. There will be a reason code portion which will accompany the anomaly score and rank output to help the users understand reason for the model output. The reason code has 3 parts: A. Behavior-level ranking B. Percentile of firm behavior C. The number of standard units away from the mean.
* The second output of firm-centric uses a statistical test to generate a small list of firms.

The goal of this project is to generate a small list of firms for the CAT Compliance team to review every month.

There is no ground truth or labels for this problem so creating the right metric to track performance of the model is difficult. Creating a label for this problem is not the right path due to the high likelihood of model degradation and constant generation of hand-created labels. In addition, I have not found a single example of current unsupervised machine learning model in production.

There are two separate outputs from firm-centric view. The first is the monthly anomaly score and rank per peer-group. The users are interested in looking at the top firms from each peer group/cluster.

There will be a reason code portion which will accompany the anomaly score and rank output to help the users understand reason for the model output. The reason code has 3 parts:

A. Behavior-level ranking

B. Percentile of firm behavior

C. The number of standard units away from the mean.

**Verification:**

Suggested method for evaluating ranks

The problem of equal weighting to feature ranks

مواردی که داخل ایمیل ازشون خواسته بودین، نتایج رو بهمون بدن

Check dashboard (UI) and API

دشبورد نیاز به چک و تأیید شما دارد.

What business metrics will be used to judge the success of this project?

Review rate of the top 20 firms per cluster. This is defined as the number of firms reviewed by users per 20 firms per cluster. ( Where are these results?????)

این نتایج موجود نیست

Does the model meet the identified design objectives? 🞏

Does the model meet the identified business purposes? 🞏

اگر اهداف حاصل شده لطفا این دو مورد رو تیک بزنین

**Data Assurance**

Provide text and/or evidence that support the following was followed as per the Model Card:

* appropriate data quality checks were performed
* feature selection process
* usage of data proxy, if applicable, was appropriate
* training/evaluation split methodology

**Summary**

All necessary data is being onboarded to MDF. The following data sources are onboarded to MDF:

* IM\_events\_summary
* Individual queries for CAT Compliance Reporting Patterns and sub-patterns
* Individual queries for Trade Reporting Patterns and sub-patterns

This is an unsupervised learning model that does not train data.

Data types are Inferred as numerical or categorical.

**Verification:**

Training/Validation data was tested to confirm the absence of nulls in the data.

این موارد نیاز به چک و تأیید شما دارد

For any data types that are inferred, how will the code handle situations where the data is missing or different from what the rest of the code expects to infer?

Categorical variable is a result of a rule-based output, so the check is done there before modeling. There is no code that explicitly handles data type mismatch for numerical data types.

Was the data approach from the Model Card Followed? 🞏

اگر اهداف حاصل شده لطفا این مورد رو تیک بزنین

**Model Selection & Training**

**Summary**

The model used is an unsupervised anomaly detection.

This is an unsupervised learning model that does not train data.

It's non-parametric which means no tuning is needed. In addition, it's easy to interpret and performs well even if the model input assumption is broken.

**Verification:**

Why were the final algorithms chosen over other alternatives considered?

Because of interpretability and rate of review.

Compare result of two other models (Isolation Tree and Histogram-Based Outlier Score (HBOS) with this model. (based on interpretability and rate of review)

Testing these three models on a dataset with labels. ( 30 benchmark datasets present in reported reference:

<https://ieeexplore.ieee.org/abstract/document/9737003>

or 55 benchmark datasets, named “ADBench” in reported reference2:

<https://arxiv.org/pdf/2206.09426.pdf>

این موارد پیشنهادات خودم هست و نیاز به تأیید شما دارد اگر اوکی نبودن لطفا پاک کنین.

Was the Model Selection approach from the Model Card Followed? 🞏

**Model Monitoring Plan**

**Summary**

N/A

**Verification:**

N/A

Is the Proposed Model Monitoring Plan satisfactory? 🗷

**Assurance Goals**

**Summary:**

The goal of this project is to generate a small list of firms for the CAT Compliance team to review every month.

What business metrics will be used to judge the success of this project?

Review rate of the top 20 firms per cluster. This is defined as the number of firms reviewed by users per 20 firms per cluster. ( Where are these results?????)

این نتایج موجود نیست

**Verification:**

N/A

Was the type and level of Assurance activities performed sufficient? 🞏

اگر اهداف حاصل شده لطفا این مورد رو تیک بزنین

**Risks**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Risk Type (Business/Technical/Other)** | **Risk Description** | **Severity** | **Mitigation** | **Potential Impact** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

**Sign Off**

Provide a formal sign off on the release of the model under test.

Click or tap here to enter text.