# **Governance & Control**

In any analytics project/undertaking, it is important to outline a Governance & Control procedure acknowledging the limitations, downsides, and risks of using the model and analytical framework put forward by the members of the project. Decision-making with the help of decision-support algorithms and models introduces certain blind-spots and biases into an organization’s decision-making framework. A Data and Model Governance policy helps ameliorate the damages arising from these side-effects, as well as helps to mitigate and prevent systematic errors from creeping into an organization’s decision criteria.

Blind-spots in data and modelling may lead to biased decisions especially when those blind-spots are related to rare events/classes. The risks arising from these blind-spots can be classified into several different categories/tiers. In this project, I will be using the Risk Tiers provided by the European Commission where risks are classified into four distinct classes, namely:

I established a four-step control procedure for this project. However, given the nature of this project – it will not be possible to establish the fourth and final step of the control procedure. The procedure consisted of:

1. Data Control: In this step, raw data is reviewed, validated, and prepped for feature engineering. Random samples of the raw data are reviewed by a human – the sample size is determined as per the requirements of the project, and size of the dataset.
2. Feature Control: After the raw data has been validated and prepped into a machine-readable/compatible format, existing features are reviewed for skewness, and missing value issues. Feature Engineering is carried out, and variable-level monitoring thresholds are established based on observation, empirical research, and domain knowledge. Variables related to sensitive information like Gender, Age, and Location are monitored or culled as these can lead to bias and fairness issues.
3. Model Control: Once feature engineering has been carried out, modelling assumptions are explicitly listed in order to detect and root out any bias and fairness issues. Appropriate models are chosen based on the analytics problem, and model validation criteria are created. Appropriate metrics to test model performance are chosen.
4. Decision Control: Model results are compared to historical benchmarks, and significant deviations in trends are closely examined in order to ensure robustness of models. Variable-level monitoring is established in order to ensure conformity with modelling environment. Industry experts will be invited to review model results, and critically examine modelling assumptions. It is important to note that these steps are iterative, and not strictly linear. Deficiencies revealed in one step will prompt the entire process to start over again.

## **A moving target**

Given the nature of the task at hand, it is important to acknowledge that the model created will have limited applicability in the future. In fraud detection, adversarial actors often change their behavior in order to bypass protective measures. Detecting fraudulent cases is therefore a ‘moving target’, and behavioral changes in the future on the part of fraudulent job posters will impact the model’s accuracy. The main goal of this project was to reveal behavioral patterns and insights into the posting characteristics of fraudulent job seekers – not to create a model that predicts fraud.

In particular - the company profile, description, benefits, and industry field characteristics may change in the future, greatly impacting the model’s prediction ability. For instance - in the current dataset, the Oil & Gas Industry has the largest number of fraudulent job posts. In the future, this may change as the Oil & Gas Industry is phased out in favor of Renewables.

In order to detect this drift/change, constant examination of recent job postings will be required. This process can be partially automated by setting up control limits for post text characteristics – such as the number of whitespaces used, number of consecutive punctuation marks used, etc. These characteristics can be monitored automatically, and any post which breaches set limits (determined empirically) can be flagged for a human auditor to review.

## **Model Inventory & Validation Plan**

## **Model Stability**

The model’s health and stability will naturally deteriorate over time as fraudsters amend their tactics. The robustness of the model can be measured from time-to-time by calculating the ROC and F1-score, as well as conducting the Kolmogorov Smirnov test. When the model’s score drops below 0.7, re-training will be prompted. At the same time, an analyst will be required to audit newer job posts to detect significant deviations from trends which existed when the model was created/trained. Additionally, if the model fails to perform better than baseline (guessing majority class) then re-training will be prompted).

## **Treatment of Nulls and Missing Values**

## **Variable Level Monitoring**

*Highlighted rows are original variables from the dataset. Non-highlighted rows are feature engineered variables.*

Table of Acceptable Value Ranges

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variable Name | Measurement Level | Number missing | Mean | Standard Deviation | Percentage of Positives (True) |
| title |  | object |  |  |  |
| location | Nominal | object |  |  |  |
| department | Nominal | object |  |  |  |
| salary\_range | Ordinal | object |  |  |  |
| company | Nominal | object |  |  |  |
| description |  | object |  |  |  |
| requirements |  | object |  |  |  |
| benefits |  | object |  |  |  |
| telecommuting | Binary | object |  |  |  |
| has\_company\_logo | Binary | object |  |  |  |
| has\_questions | Binary | object |  |  |  |
| employment\_type | Nominal | object |  |  |  |
| required\_experience | Ordinal | object |  |  |  |
| required\_education | Ordinal | object |  |  |  |
| industry | Nominal | object |  |  |  |
| function | Nominal | object |  |  |  |
| fraudulent | Binary | object |  |  | 4.8434% |
| Country | Nominal | object |  |  |  |
| City | Nominal | object |  |  |  |
| State | Nominal | object |  |  |  |
| salary\_mean | Interval | float | 657836.7202 | 22268016.4690 |  |
| company\_profile\_cleaned |  | object |  |  |  |
| requirements\_cleaned |  | object |  |  |  |
| benefits\_cleaned |  | object |  |  |  |
| description\_cleaned |  | object |  |  |  |
| commas\_desc | Ratio | int |  |  |  |
| commas\_profile | Ratio | int |  |  |  |
| commas\_requirement | Ratio | int |  |  |  |
| commas\_benefits | Ratio | int |  |  |  |
| spaces\_desc | Ratio | int |  |  |  |
| spaces\_profile | Ratio | int |  |  |  |
| spaces\_requirement | Ratio | int |  |  |  |
| spaces\_benefits | Ratio | int |  |  |  |
| company\_profile\_consecpunc | Ratio | int |  |  |  |
| requirements\_consecpunc | Ratio | int |  |  |  |
| description\_consecpunc | Ratio | int |  |  |  |
| requirements\_consecpunc | Ratio | int |  |  |  |
| benefits\_consecpunc | Ratio | int |  |  |  |
| has\_industry | Binary | Boolean |  |  |  |
| has\_function | Binary | Boolean |  |  |  |
| has\_requirements | Binary | Boolean |  |  |  |
| has\_benefits | Binary | Boolean |  |  |  |
| has\_description | Binary | Boolean |  |  |  |
| has\_employent\_type | Binary | Boolean |  |  |  |
| has\_salary | Binary | Boolean |  |  |  |
| has\_location | Binary | Boolean |  |  |  |
| company\_profile\_cleaned\_clickbait\_count | Ratio | int |  |  |  |
| requirements\_cleaned\_clickbait\_count | Ratio | int |  |  |  |
| description\_cleaned\_clickbait\_count | Ratio | int |  |  |  |
| benefits\_cleaned\_clickbait\_count | Ratio | int |  |  |  |
| company\_profile\_cleaned\_clickbait\_ratio | Ratio | float |  |  |  |
| requirements\_cleaned\_clickbait\_ratio | Ratio | float |  |  |  |
| description\_cleaned\_clickbait\_ratio | Ratio | float |  |  |  |
| benefits\_cleaned\_clickbait\_ratio | Ratio | float |  |  |  |
| company\_profile\_cleaned\_words\_per\_sentence | Ratio | float |  |  |  |
| requirements\_cleaned\_words\_per\_sentence | Ratio | float |  |  |  |
| description\_cleaned\_ words\_per\_sentence | Ratio | float |  |  |  |
| benefits\_cleaned\_ words\_per\_sentence | Ratio | float |  |  |  |
| company\_profile\_cleaned\_read\_score | Interval | float |  |  |  |
| requirements\_cleaned\_words\_ read\_score | Interval | float |  |  |  |
| description\_cleaned\_read\_score | Interval | float |  |  |  |
| benefits\_cleaned\_ read\_score | Interval | float |  |  |  |
| company\_profile\_cleaned\_sentiment | Interval | float |  |  |  |
| requirements\_cleaned\_words\_sentiment | Interval | float |  |  |  |
| description\_cleaned\_read\_sentiment | Interval | float |  |  |  |
| benefits\_cleaned\_sentiment | Interval | float |  |  |  |
| company\_profile\_external\_ref | Binary | Boolean |  |  |  |
| requirements\_cleaned\_ external\_ref | Binary | Boolean |  |  |  |
| description\_cleaned\_external\_ref | Binary | Boolean |  |  |  |
| benefits\_external\_ref | Binary | Boolean |  |  |  |

## **Risk Tiering**

The risk of falsely identifying a job post as fraudulent is High since it can lead to soured relations with job posters/employers. Failing to identify fraudulent job posts will also be detrimental to job seekers and candidates as this might lead to identify fraud, theft of personal information, etc.

|  |  |  |
| --- | --- | --- |
| Risk | Risk Level | Action |
| Model fails to detect fraudulent job post | High | * Report (as long as false negatives do not exceed 36%) * If false negatives exceed 36%, a complete overhaul of the analytical methodology has to be conducted along with model re-training. New features will have to be engineered. |
| Model falsely flags a job post as fraudulent | High | Report and communicate with job poster/employer to seek clarification |
| Model incorrectly flags a style of writing as fraudulent | Limited Risk | Model may become biased against a style of writing (personal selling, marketing). In this situation, re-training has to be conducted. |
| Incorrectly working packages | Unacceptable |  |

## **Online Analytical Processing & Continuous Analytics**