**Governance & Monitoring**

**1. Model Purpose & Business Context**

**1.1 Objective**

This model predicts the probability that an employee will experience **High Burnout**. The aim is to identify high-risk individuals early so that **proactive, supportive interventions** can be implemented. This is **not** for punitive measures but for improving well-being, retention, and productivity.

**1.2 Business Need**

* **Burnout** impacts retention, productivity, and healthcare costs.
* Remote/hybrid work, though flexible, has led to increased **social isolation** and **work-life imbalance**, which are correlated with higher burnout rates.
* A predictive system allows **HR** to focus interventions where they’re most needed, saving resources and reducing turnover.

**1.3 Executive Model Selection**

* **Random Forest (RF)** chosen for deployment:
  + **Accuracy:** 0.66
  + **Recall (High Burnout):** 0.98 → ensures most high-burnout cases are flagged.
  + **ROC AUC:** 0.56 → modest discriminative ability but balanced bias-variance trade-off.
  + **Interpretability:** Feature importance ranking identifies top drivers.

**Top Predictors:** Work Arrangement, Social Isolation Score, Work-Life Balance Score, Physical Health Issues, Salary Range.

**2. Initial Model Fit Statistics (Build vs Holdout)**

| **Model** | **Accuracy** | **ROC AUC** | **Recall (High)** | **Notes** |
| --- | --- | --- | --- | --- |
| Decision Tree | 0.59 | 0.54 | 0.39 | Easy to interpret; weaker performance. |
| **Random Forest** | **0.66** | **0.56** | **0.98** | Best for operational deployment. |
| Gradient Boosting | 0.65 | 0.52 | 0.09 | Struggles to detect minority class. |
| Logistic Regression | 0.64 | 0.54 | 0.79 | Good baseline; interpretable. |

**Interpretation:**  
Random Forest provides the highest recall for High Burnout, prioritizing sensitivity to avoid missing at-risk employees.

**3. Model Drift Monitoring – PSI (Population Stability Index)**

**3.1 Method**

* PSI compares the distribution of predicted probabilities between the **build (train)** dataset and the **holdout (test)** dataset.
* **Thresholds:**
  + <0.10 → No action
  + 0.10–0.20 → Monitor/Report
  + ≥0.20 → Retrain/Rebuild

**3.2 Result**

* **PSI = 9.7101**
* Significantly exceeds 0.20 → indicates substantial difference in score distributions between build and holdout sets.
* **Tiering:** RED → Retrain/Rebuild required before deployment.

**3.3 Implication**

The large PSI here is due to differences between training and test sample distributions; in production, this would signal either:

* Concept drift (relationship between predictors and target has changed)
* Input distribution shift (e.g., more remote workers than in training data)

**4. Variable-Level Monitoring**

**4.1 Build-Time Variable Statistics**

**Numeric Variables:**

* *Hours\_Per\_Week*: Mean ≈ 40, Median = 40, Min = 20, Max = 60, p01 = 30, p99 = 50
* *Social\_Isolation\_Score*: Mean ≈ 3.2, Median = 3, p01 = 1, p99 = 5
* *Work\_Life\_Balance\_Score*: Mean ≈ 3.1, Median = 3, p01 = 1, p99 = 5

**Categorical Variables:**

* *Work\_Arrangement*: Remote (40%), Hybrid (35%), Onsite (25%)
* *Region*: North America (45%), Asia (25%), Europe (20%), Other (10%)

**4.2 Acceptable Ranges & Missing Treatment**

* **Numeric:**
  + Range: [p01, p99]
  + Missing Imputation: **Median**
  + Out-of-range values capped to nearest boundary.
* **Categorical:**
  + Range: Valid build-time categories only
  + Missing Imputation: **Mode**
  + New categories in production → “Other/Unknown”

**Example:**

* *Mental\_Health\_Status*: ~25% missing → impute mode.
* *Physical\_Health\_Issues*: ~9% missing → impute mode.

**5. Variable Drift Examples – CSI & Chi-Square**

**5.1 Numeric Variables – CSI**

* *Hours\_Per\_Week*: CSI = 0.013 → Stable
* *Work\_Life\_Balance\_Score*: CSI = 0.007 → Stable
* *Social\_Isolation\_Score*: CSI = 0.004 → Stable

**5.2 Categorical Variables – Chi-Square**

* Variable Tested: *Work\_Arrangement*
* Chi² = 2.306, p = 0.316 → No significant shift (p ≥ 0.05)
* **Interpretation:** Categorical distribution stable.

**6. Risk Tiering Framework**

| **Metric** | **Green (No Action)** | **Amber (Watch)** | **Red (Refit/Rebuild)** |
| --- | --- | --- | --- |
| PSI / CSI | < 0.10 | 0.10–0.20 | ≥ 0.20 |
| Chi-Square p | ≥ 0.05 | — | < 0.05 |

**7. Monitoring Cadence**

**Weekly:**

* PSI on model scores
* Number and % of employees flagged as high risk
* Check for data anomalies

**Monthly:**

* CSI for numeric variables
* Chi-Square for categoricals
* Fairness review (arrangement, role, region)

**Quarterly:**

* Full model validation
* Threshold adjustment/retrain decision

**8. Fairness & Privacy**

**8.1 Fairness**

* Compare model performance metrics across:
  + Work arrangement types
  + Job roles
  + Regions
* If disparities found → adjust threshold or retrain with re-weighting.

**8.2 Privacy**

* Restrict model outputs to HR professionals with role-based access.
* Log every scoring event with user ID, date/time, and action taken.

**9. Change Control Process**

**Triggers:**

* PSI ≥ 0.20 or CSI ≥ 0.20
* Chi-Square p < 0.05 for key variable
* Policy change or new business rules

**Process:**

1. Flag breach in monitoring dashboard.
2. Retrain model using updated data.
3. Run full validation suite.
4. Document changes & rationale.
5. Approval → Model Owner → Reviewer → Approver.
6. Deploy updated model to production.

**10. Run-Book for Daily/Weekly Operations**

1. **Data Intake:** Validate schema, acceptable ranges, and impute missing values.
2. **Model Scoring:** Apply model, record predictions and top contributing features.
3. **Monitoring:**
   * Weekly PSI report
   * Monthly variable drift report
4. **Alerts:** Auto-trigger ticket if thresholds breached.
5. **Review & Action:** Investigate cause, decide retrain/rebuild.
6. **Documentation:** Update governance log with each cycle’s results.