Fair predictions of gender-sensitive recidivism

GA No: 101160473

Deliverable D4.2

System User Manual

April 2026

Document Information

Project Name	Fair predictions of gender-sensitive recidivism
Project Acronym	FAIR-PReSONS
Grant Agreement No	101160473
Call	JUST-2023-JACC-EJUSTICE
Topic	JUST-2023-JACC-EJUSTICE
Type of action	JUST Lump Sum Grants
Granting authority	European Commission-EU
Deliverable No	D4.2
Deliverable Name	System User Manual
Work Package	WP4
Dissemination Level[1]	PU– Public
Deliverable Type[2]	R – Document, report
Lead Organization	IOTAM
Lead Member	Konstantinos Kotis, Andreas Siafakas
Submission month	M24
Date	31 May 2026

[1] "PU – Public" or "SEN – Sensitive"

[2] "DATA – data sets, microdata, etc", "DEC – Websites, patent filings, videos, etc", "DEM – Demonstrator, pilot, prototype", "R — Document, report" or "OTHER"

Document history

Revision	Date	Description	Author
0.1	15/03/26	Original Draft	Andreas Siafakas
0.2	15/04/26	Partners' Feedback Integration and Finalised System Use Cases	Andreas Siafakas, Konstantinos Kotis

Authors

Partner	Name(s)
IOTAM	Andreas Siafakas
UAEGEAN	Konstantinos Kotis

Contributors

Partner	Contribution type	Author
IOTAM	Review and Feedback	Elena Valari
UAEGEAN	Review and Feedback	Stamatis Chatzistamatis
CSD	Review and Feedback	Tatyana Novossiolova
IPS	Review and Feedback	Alvaro Garcia

Glossary

Abbreviation	Description

Contents

1. INTRODUCTION 7
1.2. Purpose and Strategic Context 7
1.2. Definition of Recidivism in FAIR-PReSONS 7
1.3. Intended Audience and Use Context 8
2. Conceptual Foundations of Risk Assessment 8
2.1. The Risk Assessment Paradigm in Criminal Justice 8
2.2. Static and Dynamic Risk Factors: Conceptual Clarification 9
2.3. The Age-Crime Curve and its Judicial Implications 10
2.4. Criminal History as a Structural Predictor 10
2.5. Socio-Economic Proxies and Structural Vulnerability 11
2.6. Risk Stratification Vs Deterministic Prediction 11
2.7. Judicial Responsibility in Risk Informed Decision Making 11
3. System Architecture Overview 12
3.1. General System Design Philosophy 12
3.2. Data Pipeline and Model Training Workflow 12
3.3. Fairness-Aware Optimisation 13
3.4. Accuracy – Fairness Trade – Off 13
3.5. Knowledge Structuring and Semantic Alignment 14
3.6. Explainability Layer Engineering 14
3.7. Deployment Architecture 14
4. Accessing the Platform 15
4.1. Institutional Access Framework 15
4.2. Authentication 15
4.3. Data protection and GDPR Compliance 16
4.4. Ethical Preconditions for System use 16
4.5. Step by Step Access Procedure 17
4.6. Error Handling and Technical Limitations 17
5. Input Interface: Structure and Variables 17
5.1. Design Logic of the Input Interface 17
5.2. Demographic Variables 18

Age 18
Gender 18
Nationality 18
5.3. Criminal History Variables 18
5.4. Offence Characteristics and Crime Typologies 19
5.5. Sentencing and Institutional Variables 19
5.6. Range Constraints and Data Integrity 19
5.7. Practical Training Exercise 20
6. Greek Model 20
6.1. Overview of the Greek Dataset 20
6.2. Construction of the Recidivism Outcome Variable 20
6.3. Descriptive Patterns Observed in the Greek Dataset 21
6.4. Predictor Variables and Theoretical Justification 21
6.5. Fairness Evaluation in the Greek Model 22
6.6. Model Validation and Performance 23
6.7. Interpretation of Greek Outputs 23
6.8. Greek Use Cases – Training Scenarios 24
6.8.1. Greek Use Case 1 24
6.8.2. Greek Use Case 2 26
7. Bulgarian Model 28
7.1. Overview of the Bulgarian Dataset 28
7.2. Construction of the Outcome Variable 29
7.3. Descriptive Patterns in the Bulgarian Dataset 29
7.4. Predictor Variables 30
7.5. Risks Assessment Category Variable 31
7.6. Fairness Evaluation in the Bulgarian Context 31
7.7 Model Validation 32
7.8. Interpretation of Bulgarian Outputs in Judicial Context 32
7.9. Bulgarian Use Case 1 33
7.10. Bulgaria Use Case 2 35
8. Portuguese Model 39

8.1. Overview of the Portuguese Dataset and Modelling Context.. 39
8.2. Construction of the Recidivism Risk Framework.. 39
8.3. Risk Factor Variables...40
8.4. Fairness Evaluation in the Portuguese Model.. 42
8.5. Model Validation.. 42
Portugal Use Case 1..43
Portugal Use Case 2..45

1. INTRODUCTION

1.2. Purpose and Strategic Context

This User Manual accompanies the FAIR-PReSONS decision-support platform and is designed for use during national judicial trainings in Greece, Bulgaria, and Portugal. It serves as an operational, methodological, and ethical guide for judicial professionals interacting with the AI-assisted recidivism risk assessment system.

The manual has three primary objectives:

1. To explain the conceptual and methodological foundations of the system.
2. To provide detailed operational guidance on system use.
3. To support critical, legally grounded interpretation of model outputs.

FAIR-PReSONS was developed within the framework of the EU Justice Programme to respond to longstanding concerns about opacity, bias, and lack of contextual adaptation in algorithmic recidivism tools. Unlike off-the-shelf predictive systems developed in other jurisdictions, FAIR-PReSONS has been constructed using national datasets, fairness-aware optimisation techniques, and explicit legal-ethical constraints aligned with European standards.

The system is explicitly positioned as a decision-support tool. It does not replace judicial reasoning, nor does it produce binding recommendations. Its outputs must always be interpreted in conjunction with case files, contextual factors, and principles of proportionality and individualisation.

1.2. Definition of Recidivism in FAIR-PReSONS

Within FAIR-PReSONS, recidivism is defined specifically as reincarceration following release from prison. This definition excludes rearrests that do not result in custody, technical probation violations (unless leading to custodial admission), and purely administrative returns.

The operationalisation differs slightly across national contexts due to dataset architecture:

- In Greece, recidivism is defined as reincarceration due to a new offence within a defined post-release time window (typically three years).
- In Bulgaria, due to data constraints, an event-based proxy measure is constructed distinguishing custodial admission from probation admission.

- In Portugal, recidivism risk is estimated through structured administrative variables aligned with validated criminological predictors.

These methodological differences reflect responsible adaptation to national institutional realities rather than conceptual divergence.

1.3. Intended Audience and Use Context

This manual is intended for:

- **Judges and judicial trainees;**
- **Prosecutors and court clerks;**
- **Probation officers and correctional administrators;**
- **Policy analysts participating in pilot implementation.**

The document is structured to support workshop-based training, scenario simulation, and structured case interpretation exercises.

2. Conceptual Foundations of Risk Assessment

2.1. The Risk Assessment Paradigm in Criminal Justice

Risk assessment has become a central component of contemporary criminal justice systems. Since the 1990s, criminological research has increasingly shifted from purely retributive sentencing philosophies toward structured approaches that incorporate predictive indicators of future offending. This development is closely associated with the emergence of the Risk–Need–Responsivity (RNR) framework and related meta-analytical research identifying statistically robust predictors of recidivism.

Within this paradigm, recidivism prediction is not understood as prophecy but as probabilistic risk stratification. Courts are not attempting to foresee individual destiny; rather, they evaluate whether a given individual shares characteristics with historical cohorts that exhibited higher or lower rates of reoffending.

It is crucial to distinguish between three levels of reasoning:

1. **Descriptive analysis** – identifying statistical correlations within historical datasets;
2. **Predictive modelling** – estimating likelihood based on structured inputs;

3. **Normative judicial reasoning** – determining appropriate legal response within constitutional constraints.

FAIR-PReSONS operates exclusively at the second level. It translates structured case attributes into probabilistic estimates derived from historical data. It does not engage in normative judgment. That responsibility remains entirely with the judiciary.

2.2. Static and Dynamic Risk Factors: Conceptual Clarification

Criminological literature distinguishes between static and dynamic risk factors.

Static Factors

Static factors refer to variables that are historically fixed or not meaningfully modifiable in the short term. These include:

- Age at first conviction or incarceration;
- Number of prior convictions or incarcerations;
- Gender;
- Criminal history patterns;
- Certain structural offence characteristics.

Static predictors consistently demonstrate strong statistical association with recidivism. Age, for instance, follows the well-established "age–crime curve," where criminal involvement peaks in late adolescence and early adulthood and declines with age. Prior convictions are among the most powerful predictors of future criminal justice contact.

The advantage of static variables lies in their objectivity and reproducibility. They reduce subjective interpretation and are reliably documented in administrative datasets.

Dynamic Factors

Dynamic factors refer to variables that can change over time and may be influenced by intervention. Examples include:

- Employment status;
- Educational attainment;
- Substance abuse treatment participation;
- Institutional behaviour;

- Participation in rehabilitation programmes.

Dynamic factors are highly relevant for rehabilitation-oriented policy. However, they are often inconsistently recorded across national administrative systems. For this reason, FAIR-PReSONS relies primarily on static predictors supplemented by limited proxy indicators for socio-economic vulnerability.

Judicial users must therefore understand that the system reflects structural and historical patterns more strongly than recent behavioural transformation.

2.3. The Age-Crime Curve and its Judicial Implications

The relationship between age and criminal behaviour is one of the most robust findings in criminology. Empirical studies across jurisdictions demonstrate that criminal activity rises sharply during adolescence, peaks in early adulthood, and gradually declines thereafter.

This pattern has multiple explanations:

- Neurodevelopmental maturity;
- Social integration through employment and family formation;
- Reduced peer-driven risk behaviour;
- Institutional deterrence effects.

Within the FAIR-PReSONS models, age is encoded categorically to allow non-linear effects. Younger age bands may increase estimated probability not because of moral evaluation but because historical cohorts in that range exhibited higher rates of reincarceration.

Judges must interpret this carefully. Age increases statistical risk in the model, but it also interacts with legal principles such as diminished culpability for younger individuals and heightened rehabilitation potential.

2.4. Criminal History as a Structural Predictor

The number of prior incarcerations is a consistently strong predictor across all three national datasets. Individuals with repeated custodial episodes statistically display higher probabilities of future custodial return.

From a modelling perspective, this variable captures persistence and institutional cycling patterns. From a judicial perspective, it must be interpreted in light of proportionality and the principle of individualisation. Repeated incarceration may reflect structural disadvantage, addiction cycles, or limited access to reintegration support.

The model quantifies recurrence patterns; it does not attribute moral blame.

2.5. Socio-Economic Proxies and Structural Vulnerability

Educational attainment and employment status appear in all three national contexts as relevant predictors. Lower levels of education and unemployment at entry correlate with higher reincarceration rates.

These variables function as proxies for socio-economic vulnerability. They do not imply causation but reflect structural barriers to reintegration.

The inclusion of such variables raises important ethical questions. Could modelling socio-economic disadvantage inadvertently reinforce inequality? FAIR-PReSONS addresses this concern through fairness-aware optimisation, ensuring that inclusion of predictive variables does not translate into disproportionate error rates across protected groups.

Judicial users should recognise that socio-economic variables highlight vulnerability and may inform targeted support measures rather than punitive intensification.

2.6. Risk Stratification Vs Deterministic Prediction

A fundamental distinction must be maintained between risk stratification and deterministic forecasting.

The model estimates probability based on similarity to historical data. It does not claim that a given individual will reoffend. A 65% estimated probability does not mean that the person will reoffend with certainty; it means that among historically similar individuals, approximately 65% were reincarcerated within the defined window.

Judges should therefore interpret outputs as conditional statistical estimates. Risk categories are comparative tools, not declarations of future fact.

2.7. Judicial Responsibility in Risk Informed Decision Making

The integration of predictive tools into judicial environments introduces both opportunity and responsibility.

Opportunities include:

- Increased consistency in structured reasoning;
- Awareness of empirically validated predictors;
- Identification of high-risk profiles requiring targeted intervention.

Responsibilities include:

- Ensuring proportionality;
- Avoiding automation bias;
- Recognising limitations of administrative data;
- Preserving individualised assessment.

The FAIR-PReSONS system is therefore best understood as an analytical instrument within a broader legal framework. Its outputs must be critically evaluated, contextualised, and integrated into constitutional reasoning.

3. System Architecture Overview

3.1. General System Design Philosophy

The FAIR-PReSONS platform has been designed as a modular, fairness-aware, human-centred decision-support system. Its architecture reflects three foundational principles:

1. **Contextual Adaptation** – Each national model is trained independently using country-specific datasets.
2. **Fairness-by-Design** – Bias mitigation mechanisms are embedded directly into the model training process.
3. **Explainability and Auditability** – All predictions are accompanied by interpretable feature contribution outputs.

The system does not consist of a single pan-European predictive engine. Instead, it contains three parallel national models (Greece, Bulgaria, Portugal), each developed according to national data structures, institutional realities, and legal frameworks. This prevents inappropriate cross-country generalisation and preserves contextual integrity.

3.2. Data Pipeline and Model Training Workflow

The development of each national model followed a structured pipeline:

1. **Data Collection and Anonymisation** – Administrative prison datasets were collected in collaboration with national authorities. Personally identifiable information was removed prior to modelling.
2. **Data Cleaning and Preprocessing** – Missing values, inconsistent encodings, and structural irregularities were addressed. Variables were harmonised into structured formats suitable for machine learning.
3. **Outcome Variable Construction** – Each country operationalised recidivism according to its dataset architecture.
4. **Feature Engineering** – Variables were encoded categorically or numerically depending on structure. Age bands, crime typologies, and institutional indicators were harmonised.
5. **Model Training with Fairness Constraints** – Algorithms were trained under dual optimisation goals: predictive performance and fairness parity.
6. **Validation and Testing** – Performance and fairness metrics were evaluated across multiple data splits.

3.3. Fairness-Aware Optimisation

A defining characteristic of FAIR-PReSONS is the integration of fairness constraints directly into the optimisation objective of the machine learning models.

Traditional predictive systems optimise solely for accuracy, minimising prediction error across the entire dataset. However, such optimisation may produce unequal error rates across demographic groups.

FAIR-PReSONS incorporates fairness metrics—such as:

- **Equal Opportunity Difference** (minimising disparity in false negative rates);
- **Equalized Odds** (balancing both false positive and false negative disparities);
- **Disparate Impact Ratio** (monitoring proportionality of positive classifications);

These metrics are encoded into the model's loss function as additional penalty terms. During training, the algorithm simultaneously minimises predictive error and disparity across protected attributes.

This approach differs fundamentally from post-hoc fairness correction. Instead of adjusting outputs after training, fairness is embedded into the model's internal learning process.

3.4. Accuracy – Fairness Trade – Off

In fairness-aware machine learning, a central challenge is balancing predictive accuracy with equity constraints.

Maximising accuracy alone may produce group disparities. Imposing strict parity constraints may reduce overall predictive performance.

FAIR-PReSONS adopts a calibrated optimisation approach, seeking a balanced equilibrium where:

- Predictive performance remains within acceptable statistical thresholds;
- Error disparities across groups are meaningfully reduced;
- No protected group systematically bears disproportionate false negatives or false positives.

Judicial users should understand that the model's outputs reflect this balanced optimisation. The system does not pursue maximal predictive power at the expense of fairness.

3.5. Knowledge Structuring and Semantic Alignment

Beyond numerical modelling, FAIR-PReSONS integrates structured data alignment principles.

Variables were not treated as isolated numeric entries but as semantically meaningful legal attributes. Crime categories, sentence types, and release modalities were harmonised to reflect legal classifications rather than arbitrary codes.

This semantic structuring supports:

- Interpretability;
- Legal traceability;
- Potential interoperability with future judicial information systems.

The modelling architecture therefore operates at the intersection of statistical computation and legal taxonomy.

3.6. Explainability Layer Engineering

Explainability is not an optional feature but a mandatory component of high-risk AI systems within the European regulatory landscape.

FAIR-PReSONS incorporates feature contribution analysis (e.g., SHAP-based visualisations) that decomposes each prediction into variable-level contributions relative to baseline population risk.

This allows users to observe:

- Which factors increased the predicted probability;

- Which factors reduced it;
- The relative magnitude of each contribution.

Explainability serves three purposes:

1. **Transparency** – Users can understand how the model reached its output.
2. **Contestability** – Outputs can be challenged or scrutinised.
3. **Educational Value** – Judicial professionals can observe empirical relationships embedded in the data.

3.7. Deployment Architecture

The operational platform is deployed as a web-based interface with structured input fields and real-time output generation.
Importantly:

- The system does not automatically store case-level operational data;
- It does not autonomously communicate decisions to external systems;
- It does not trigger procedural consequences.

Human oversight remains structurally embedded. Judges input case attributes voluntarily and interpret outputs independently.

4. Accessing the Platform

4.1. Institutional Access Framework

The FAIR-PReSONS platform is designed to operate within controlled judicial or training environments. Access is structured to reflect the sensitivity of high-risk AI systems operating in the criminal justice domain. Although the training prototype may operate in a simplified web-based environment, full institutional deployment assumes a governance structure involving:

- Designated institutional administrators;
- Clearly defined user roles;
- Logged access sessions;
- Secure hosting infrastructure.

In the context of national trainings, access is typically provided through a secure URL hosted by the responsible technical partner. Users do not install software locally. This minimises cybersecurity risks and ensures version control across training sessions.

The platform architecture does not require integration with live prison databases during training use. Instead, structured case attributes are entered manually by the judicial user.

4.2. Authentication

In operational deployment scenarios, user authentication mechanisms may include institutional credentials or controlled login access. Even in training mode, however, the ethical responsibility attached to system use remains identical.

Judicial professionals must understand that:

- The system processes sensitive criminal justice variables;
- Outputs may influence structured reasoning during hearings;
- Misuse or misinterpretation may generate inappropriate judicial reliance.

Therefore, access to the platform presupposes completion of introductory training modules explaining its methodological scope and limitations.

4.3. Data protection and GDPR Compliance

The FAIR-PReSONS system was developed in strict compliance with European data protection principles. During the model development phase:

- All datasets were anonymised prior to modelling;
- Personally identifiable information (names, identification numbers, addresses) was removed;
- Unique identifiers were replaced with pseudonymous codes;
- Data minimisation principles were applied.

During operational use, the platform does not require entry of personally identifiable data. Users input structured attributes (e.g., age band, crime type, sentence length) rather than names or case numbers.

This design reflects GDPR core principles:

- **Purpose limitation** – The system is used exclusively for structured risk assessment within judicial reasoning.
- **Data minimisation** – Only necessary attributes are entered.
- **Storage limitation** – The training prototype does not persistently store case-level data.
- **Integrity and confidentiality** – Secure hosting prevents unauthorised access.

Judges must avoid entering unnecessary personal details into open text fields (if any). The system does not require narrative case descriptions.

4.4. Ethical Preconditions for System use

<u>Before using the system, judicial professionals should explicitly acknowledge three ethical preconditions:</u>

1. <u>The output is probabilistic and not determinative.</u>
2. <u>The model reflects historical administrative data patterns.</u>
3. <u>Judicial discretion remains primary.</u>

These preconditions should be reiterated at the beginning of training workshops to prevent automation bias—a cognitive tendency to over-rely on algorithmic outputs.

4.5. Step by Step Access Procedure

In a standard training scenario, the workflow unfolds as follows:

1. **User navigates to the secure platform URL.**
2. **Country model is selected (Greece, Bulgaria, Portugal).**
3. **Input interface loads dynamically according to national variable structure.**
4. **Structured case attributes are entered.**
5. **User submits the case for prediction.**
6. **Results dashboard displays classification, probability, and explanation.**

4.6. Error Handling and Technical Limitations

During use, users may encounter:

- Missing field warnings;
- Invalid category errors;
- Numerical range restrictions (e.g., age bounds);
- Session timeouts.

Such safeguards are intentionally embedded to preserve model integrity. For example, age must fall within the range used during training (e.g., 15–82 in the Greek dataset). Entering out-of-range values would produce unreliable outputs.

Technical errors do not imply conceptual system failure but typically reflect data-entry inconsistencies.

5. Input Interface: Structure and Variables

5.1. Design Logic of the Input Interface

The input interface translates structured legal and administrative attributes into machine-readable variables. It is organised into thematic sections to mirror judicial reasoning categories:

- Demographics;
- Criminal history;
- Offence characteristics;
- Sentencing attributes;
- Institutional trajectory indicators.

This organisation ensures cognitive alignment between judicial case assessment and algorithmic input.

The interface intentionally avoids open-ended narrative text boxes. All variables are categorical or numerical with defined ranges. This reduces interpretative ambiguity and ensures consistency with the model's training distribution.

5.2. Demographic Variables

Demographic variables typically include age, gender, and nationality.

Age

Age may be entered as a numeric value (Greece, Bulgaria) or categorical band (Portugal). Age functions as a statistical predictor reflecting historical recidivism patterns across life stages.
Judicial interpretation must recognise that age increases or decreases estimated probability due to population-level patterns, not moral evaluation.

Gender

Gender is included as a binary variable reflecting dataset structure. In all three national datasets, male individuals represent the majority of custodial admissions.
However, fairness-aware optimisation ensures that inclusion of gender does not produce disproportionate error rates.

Nationality

Nationality appears in Greek and Portuguese models and reflects observed structural differences in reincarceration patterns. Its inclusion enables detection of systemic disparities but must be interpreted cautiously to avoid reinforcing discriminatory assumptions.

5.3. Criminal History Variables

Criminal history is among the strongest predictors in recidivism modelling.
In the Greek model, number of prior incarcerations directly captures persistence patterns. In Bulgaria, institutional trajectory variables serve a similar function. In Portugal, previous incarceration context is reflected indirectly through structured administrative categories.
Judicial users should recognise that criminal history variables represent empirical recurrence patterns, not normative judgments.

5.4. Offence Characteristics and Crime Typologies

Crime categories are harmonised into structured typologies rather than detailed legal articles. This harmonisation supports statistical stability.
For example:

- Property offences;
- Drug-related offences;
- Violent crimes;
- Crimes against public order.

Certain typologies historically exhibit higher recurrence rates. The model encodes these statistical relationships.
Judges should consider whether the typological classification appropriately reflects the case at hand before submission.

5.5. Sentencing and Institutional Variables

Sentence length, release type, custodial complexity, and institutional trajectory indicators introduce contextual nuance.
For example:

- Conditional early release may statistically correlate with different recurrence patterns compared to full sentence completion.
- Longer sentence lengths may reflect offence severity but may also interact with reintegration barriers.

These variables should be entered precisely as recorded in the official case file.

5.6. Range Constraints and Data Integrity

Numerical fields include built-in constraints to ensure alignment with training data.
For example:

- Age bounds reflect observed dataset ranges.
- Days in prison must be non-negative integers.
- Sentence length categories must match predefined bands.

Incorrect categorisation may distort prediction.

5.7. Practical Training Exercise

During workshops, facilitators should conduct a guided exercise in which:

1. **A case profile is read aloud.**
2. **Participants collectively agree on input categories.**
3. **The system is run.**
4. **Results are discussed.**

This interactive process strengthens understanding of how structured attributes translate into predictive output.

6. Greek Model

6.1. Overview of the Greek Dataset

The Greek national model was developed using a large-scale anonymised administrative dataset covering prison releases between 2018 and 2023. The dataset includes more than 42,000 release records corresponding to over 35,000 unique individuals. The availability of a persistent anonymised identifier allowed longitudinal reconstruction of incarceration trajectories.

The dataset contains demographic attributes, offence typologies, sentence characteristics, release modalities, and selected socio-economic indicators. Although certain dynamic behavioural variables (e.g., in-prison disciplinary infractions, psychological assessments, programme participation) were not consistently available, the dataset provides strong coverage of static and structural predictors.

The Greek dataset reflects real institutional practices within the Hellenic correctional system. It includes both male and female incarcerated populations, though male individuals constitute the majority of custodial admissions. This demographic imbalance reflects broader European incarceration patterns and is taken into account during fairness evaluation.

6.2. Construction of the Recidivism Outcome Variable

The primary outcome variable in the Greek model captures reincarceration due to the commission of a new offence within a defined post-release window (typically three years).

<u>The construction process involved the following methodological steps:</u>

1. Identification of release events for each anonymised individual.
2. Longitudinal tracking of subsequent incarceration entries.
3. Exclusion of returns attributable to procedural or administrative causes (e.g., appeal processing, technical parole violations, pending case adjustments).
4. Inclusion only of re-entries clearly associated with new criminal offences.

<u>The final binary variable distinguishes:</u>

- 1 = Reincarceration due to a new offence within the defined window;
- 0 = No reincarceration within the defined window.

This operationalisation ensures conceptual coherence with criminological definitions of recidivism and avoids conflating administrative cycling with substantive criminal relapse.

From a judicial perspective, this distinction is crucial. The model predicts probability of new-offence-driven custodial return, not procedural instability.

6.3. Descriptive Patterns Observed in the Greek Dataset

Before model training, extensive descriptive analysis was conducted.

Key observations included:

- Individuals under 35 at time of incarceration exhibited significantly higher reincarceration rates.
- Repeated incarceration history strongly correlated with future custodial return.
- Property and drug-related offences displayed higher recurrence rates compared to homicide or severe violent crimes.
- Lower educational attainment was overrepresented among reincarcerated individuals.
- Unemployment at entry correlated with higher reincarceration probability.

These descriptive patterns provided empirical grounding for predictor selection.

6.4. Predictor Variables and Theoretical Justification

Number of Prior Incarcerations

This variable functions as the strongest structural predictor. It captures persistence patterns and cumulative exposure to the penal system. International literature consistently identifies prior criminal justice involvement as a dominant predictor of future recidivism.

From a modelling standpoint, this variable contributes substantial explanatory power. From a judicial standpoint, it must be interpreted alongside proportionality principles.

Age at Entry

Age was encoded into categorical bands to allow non-linear modelling of risk. Younger age groups exhibit higher statistical recurrence. This aligns with the age–crime curve.

Judges should note that while younger age increases statistical probability, it simultaneously corresponds to heightened rehabilitation potential.

Gender

Gender is retained as a binary predictor due to strong empirical signal. Male individuals demonstrate higher reincarceration rates in the dataset.However, fairness constraints were explicitly applied to ensure that gender inclusion does not produce disproportionate false negative or false positive disparities.

Crime Typology

Offences were harmonised into manageable categories. Property and drug-related offences display higher recurrence rates compared to violent offences with longer custodial exposure. This reflects opportunity structures and behavioural recurrence patterns rather than moral weighting.

Nationality

Nationality displayed observable differences in reincarceration rates. Its inclusion allows the model to detect structural patterns potentially related to reintegration barriers. During fairness evaluation, particular attention was paid to ensuring that nationality does not produce disproportionate predictive harm.

Education and Employment

Educational attainment and employment status function as proxy indicators of socio-economic vulnerability. Lower education and unemployment correlate with higher reincarceration probability.

These variables highlight structural disadvantage rather than intrinsic risk.

Sentence Length and Release Type

Sentence length and release modality capture institutional trajectory context. Conditional release may interact differently with recurrence patterns compared to full sentence completion.

These variables add nuance beyond purely demographic predictors.

6.5. Fairness Evaluation in the Greek Model

During model training, fairness diagnostics were conducted across protected attributes such as gender and age.

Metrics evaluated included:

- False negative rate disparity;
- False positive rate disparity;
- Equal opportunity difference;
- Overall accuracy and F1 score.

Where disparities exceeded acceptable thresholds, fairness penalties were introduced into the loss function to rebalance optimisation.

6.6. Model Validation and Performance

The Greek model underwent cross-validation across multiple train–test splits. Performance metrics were evaluated to ensure statistical stability.
While exact performance figures are technical in nature, the model achieved balanced predictive accuracy while maintaining fairness parity within defined tolerances. Judicial users should understand that predictive performance is not perfect. No model achieves 100% accuracy. Outputs reflect probabilistic approximation.

6.7. Interpretation of Greek Outputs

The Greek model produces:

- **A qualitative risk classification (LOW / MEDIUM / HIGH);**
- **A probability estimate;**
- **A feature contribution explanation.**

Judges should interpret classification as relative position within the historical distribution.
A HIGH classification indicates that the individual shares multiple characteristics with historically higher-recurrence cohorts. It does not imply inevitability.

Feature contribution diagrams allow judicial users to observe which variables most strongly influenced the estimate. For example:

- **Multiple prior incarcerations may increase probability;**

- **Older age may reduce it;**
- **Stable employment may reduce it.**

Judges should assess whether the statistical reasoning aligns with contextual knowledge not captured in structured data (e.g., recent rehabilitation efforts).

<u>The Greek model does not incorporate:</u>

- **Detailed behavioural assessments;**
- **Psychological evaluations;**
- **Programme participation data;**
- **Post-release support quality.**

Therefore, it captures structural historical patterns rather than dynamic behavioural transformation. Judges must integrate qualitative case insights to complement the statistical output.

6.8. Greek Use Cases – Training Scenarios

The following two cases were executed in the Greek national model during training simulations. The outputs presented reflect actual system predictions and are used to illustrate interpretation methodology in practice.

7. Bulgarian Model

7.1. Overview of the Bulgarian Dataset

The Bulgarian national model was developed using an extensive administrative dataset of penitentiary admissions covering the period January 2018 to November 2024. The dataset contains over 76,000 admission records corresponding to individuals with prior incarceration experience.

Unlike the Greek dataset, the Bulgarian data architecture does not include a persistent anonymised identifier allowing precise longitudinal tracking of the same individual across multiple incarceration episodes. Each admission is recorded under a separate file code, even if it concerns the same person.

This structural characteristic significantly influenced the modelling strategy and required careful methodological adaptation in order to construct a meaningful and policy-relevant outcome variable. The dataset includes:

- Demographic information (age at entry, age at exit, gender, nationality);
- Educational level;
- Marital status and selected family indicators;
- Institutional variables (days in prison, prison status);
- Sentencing attributes;

- Release-related indicators (e.g., conditional early release, revocation);
- Risk assessment category (low / medium / high, where available).

This relatively rich institutional coverage allows for modelling of structural prison trajectory patterns, even in the absence of longitudinal identifiers.

7.2. Construction of the Outcome Variable

Due to the absence of a unique identifier enabling reconstruction of incarceration histories, a classical recidivism flag (return to prison after release) could not be constructed with precision.
To address this constraint, the modelling team developed an event-based proxy outcome variable based on the type of institutional admission.

The binary target variable is defined as:

- 1 = Admission to a correctional institution (custodial sanction);
- 0 = Admission to probation supervision (non-custodial sanction).
-

Pre-trial detention without resulting conviction was excluded from the outcome construction to preserve substantive penal meaning.

This operationalisation does not measure individual-level repeat offending in a strict longitudinal sense. Rather, it distinguishes between pathways resulting in custodial reincarceration and those diverted into supervised community sanctions.

From a policy perspective, this distinction is highly relevant. It reflects systemic re-absorption into carceral infrastructure versus diversion into reintegration pathways.

Judicial users must clearly understand that the Bulgarian model predicts probability of custodial pathway re-entry under this proxy definition—not necessarily commission of a new offence in isolation.

7.3. Descriptive Patterns in the Bulgarian Dataset

Pre-modelling descriptive analysis revealed several important structural patterns:

- Individuals aged between approximately 34 and 48 account for over half of all custodial admissions.
- Younger working-age cohorts display higher relative custodial entry rates compared to older cohorts.
- Male individuals represent approximately 96% of admissions, reflecting a significant gender imbalance in incarceration.
- Low educational attainment is highly prevalent among admitted individuals, with limited representation of tertiary education.

- A substantial proportion of admissions involve short custodial sentences (under one year), often associated with repetitive lower-level offences.

7.4. Predictor Variables

The Bulgarian model incorporates a broad set of predictors reflecting demographic, socio-economic, and institutional characteristics.

Age at Entry and Age at Exit

Age functions as a structural demographic predictor. As observed internationally, younger individuals display higher custodial re-entry likelihood.

Age at exit may capture maturity effects and time exposed to institutional intervention.

Gender

Gender displays a strong empirical imbalance in the dataset. Although male individuals dominate admissions numerically, fairness-aware modelling ensures that inclusion of gender does not generate disproportionate predictive harm.

Educational Level

Educational attainment serves as a socio-economic vulnerability proxy. Lower educational levels correlate with higher custodial admission likelihood.

This variable reflects structural disadvantage rather than intrinsic risk.

Employment and Marital Indicators

Employment status (where available) and marital status introduce dimensions of social stability. Stable employment and family integration may statistically correlate with reduced custodial return.

These indicators must be interpreted as contextual rather than moral attributes.

Institutional Trajectory Variables

The Bulgarian dataset uniquely includes several institutional trajectory variables, such as:

- Sentence fulfilled;
- Conditional early release;
- Revocation of release;
- Exemption from serving sentence;
- Reducing punishment with work;
- Interruption of implementation.

These variables reflect interactions between individuals and correctional decision processes.

For example:

- Revocation of release may statistically correlate with future custodial admission.
- Conditional early release may reflect differential supervision pathways.

Such variables capture systemic patterns rather than purely individual behavioural characteristics.

7.5. Risks Assessment Category Variable

The dataset includes an internal institutional risk assessment classification (low / medium / high) for a subset of individuals. Although coverage is incomplete, this variable introduces a unique modelling dimension.

Its inclusion allows the model to examine whether institutional risk categorisation aligns with actual custodial re-entry patterns.

However, the use of such a variable requires caution. Incorporating an existing risk classification into a predictive model risks circularity if not carefully evaluated. Therefore, fairness diagnostics were conducted to ensure that this variable does not disproportionately amplify institutional bias.

7.6. Fairness Evaluation in the Bulgarian Context

Given the gender imbalance in the dataset and the proxy nature of the outcome variable, fairness evaluation was particularly important in the Bulgarian model.

Diagnostics included:

- False negative rate comparison across gender groups;
- False positive rate comparison;
- Evaluation of predictive parity across educational strata.

Where disparities emerged, optimisation penalties were introduced to rebalance the model.

The objective was to prevent structural amplification of institutional inequalities while maintaining predictive reliability.

7.7 Model Validation

The Bulgarian model underwent repeated validation across stratified data splits to ensure stability despite the absence of persistent identifiers.

Although the proxy outcome introduces methodological limitation, performance metrics indicated consistent predictive behaviour across sub-samples.

Judicial users must recognise that predictive certainty is bounded by dataset architecture.

7.8. Interpretation of Bulgarian Outputs in Judicial Context

The Bulgarian model produces:

- Risk classification (low / medium / high);
- Probability percentage;
- Feature contribution explanation.

Because the target variable reflects custodial pathway re-entry, judicial interpretation must remain cautious.

A HIGH classification indicates statistical similarity to individuals historically admitted to correctional institutions rather than probation.

Feature contribution diagrams allow judges to observe whether institutional trajectory variables (e.g., revocation history) strongly influenced prediction.

Judges should critically assess whether structural system interaction patterns are driving risk estimates.

Key limitations include:

- Absence of persistent individual identifiers;
- Proxy outcome definition;
- Partial coverage of institutional risk assessment category;
- Limited dynamic behavioural data.

These limitations do not invalidate predictive utility but require transparent acknowledgment during interpretation.

8. Portuguese Model

8.1. Overview of the Portuguese Dataset and Modelling Context

Compared to the Greek dataset (which allowed longitudinal tracking through persistent identifiers) and the Bulgarian dataset (which required a proxy outcome due to architectural constraints), the Portuguese dataset occupies an intermediate position: it provides structured individual-level information but with certain limitations in granularity and dynamic behavioural indicators.

The modelling strategy in the Portuguese case therefore emphasised harmonisation, categorical encoding, and alignment with internationally validated criminological predictors. Rather than relying on highly granular behavioural variables, the Portuguese model draws predictive strength from carefully structured static and institutional attributes.

The dataset includes:

- Gender;
- Nationality;
- Educational level;
- Age group at release;
- Sentence length band;
- Crime category and specific offence type;
- Establishment management complexity;
- Reason for release.

Although variables such as psychological assessment scores or programme participation were not consistently available, the dataset provides strong structural indicators of reintegration context and offence patterns.

8.2. Construction of the Recidivism Risk Framework

The Portuguese model aligns closely with established criminological research identifying key predictors of reincarceration. Rather than constructing an entirely novel predictor structure, the modelling team cross-referenced national data availability with internationally recognised risk factors.

The modelling process involved:

1. Mapping available administrative variables to established predictor categories (e.g., age, offence type, socio-economic proxy);

2. Harmonising categorical encodings across datasets;
3. Structuring age and sentence variables into meaningful bands;
4. Applying fairness-aware optimisation during training;
5. Evaluating performance across risk horizons.

The resulting model estimates probability of reincarceration and, where configured, provides distinct projections for medium-term horizons (e.g., three years).

This structured alignment ensures theoretical grounding while respecting dataset constraints.

8.3. Risk Factor Variables

Age Band Encoding and Non-Linear Risk Patterns

In the Portuguese model, age is encoded into discrete bands (e.g., 18–24, 25–34, 35–44, 45–54, 55+). This approach reflects two considerations:

1. The empirical non-linearity of the age–crime curve;
2. The need for stable categorical modelling given dataset distribution.

Younger age bands historically exhibit higher rates of reincarceration. However, the model does not assume linear decline. Instead, it captures structured cohort-based differences.

From a judicial perspective, this categorical approach enhances interpretability. Judges can clearly see how belonging to a specific age band influences probability without relying on abstract numerical coefficients.

Educational Level and Socio-Economic Proxy Indicators

Educational level is encoded categorically (e.g., primary, secondary, higher education). Lower levels of educational attainment correlate with higher reincarceration rates in descriptive analysis.

This relationship reflects structural reintegration barriers rather than intrinsic behavioural risk. Lower educational attainment may limit employment opportunities and increase vulnerability to reoffending in economically motivated crime categories. The model includes this variable as a proxy indicator of socio-economic stability while fairness constraints mitigate potential amplification of structural inequality.

Crime Category and Offence Typology Harmonisation

Crime types are harmonised into broader categories to ensure statistical stability and interpretability. Rather than modelling individual penal code articles, offences are grouped into typologies such as:

- Property offences;

- Drug-related offences;
- Violent crimes;
- Public order offences;
- Other criminal acts.

This harmonisation ensures that the model captures structural recurrence patterns without overfitting to rare offence types.

Descriptive analysis indicated that property and drug-related offences exhibit higher recurrence rates compared to certain violent offences, particularly where longer custodial exposure is involved.

Judicial users should ensure that the selected typology accurately reflects the case classification before submitting inputs.

Sentence Length Bands and Institutional Exposure

Sentence length is encoded into categorical bands (e.g., less than 1 year, 1–3 years, 3–5 years, etc.).

Shorter sentences are statistically associated with higher admission frequency and may reflect lower-level but repetitive offending patterns. Longer sentences are less frequent and often correspond to more severe offences. However, sentence length interacts with reintegration dynamics. Longer incarceration may also produce institutionalisation effects or labour market exclusion, complicating simple linear interpretation. The model captures these statistical relationships without assigning normative weight to severity.

Establishment Management Complexity

A distinctive feature of the Portuguese dataset is the inclusion of establishment management complexity. This variable reflects the security level and organisational structure of the custodial environment. Complexity levels may indirectly correlate with offence severity, supervision intensity, and access to reintegration services. Including this variable introduces contextual nuance beyond purely individual-level characteristics. It acknowledges that institutional environment may shape post-release trajectories. From a judicial standpoint, this reinforces the principle that risk is influenced by systemic conditions, not solely by individual attributes.

Reason for Release

Reason for release (e.g., full sentence completion, conditional release, parole, administrative decision) is included as a predictive variable.

Different release modalities may correlate with differential supervision intensity and reintegration preparation. For example, structured conditional release may provide greater support mechanisms than abrupt full-term completion.

The model captures statistical associations between release type and reincarceration patterns. However, this variable must be interpreted carefully, as release modality may reflect prior judicial discretion.

8.4. Fairness Evaluation in the Portuguese Model

Fairness diagnostics were conducted across protected attributes such as gender and nationality.

Metrics examined included:

- False negative rate disparity;
- False positive rate disparity;
- Equal opportunity difference;
- Overall predictive performance.

Where disparities emerged, fairness penalties were incorporated into the optimisation objective to rebalance predictive behaviour.

This ensures that demographic attributes do not generate disproportionate predictive harm.

8.5. Model Validation

The Portuguese interface may provide separate outputs for:

- Overall recidivism probability;
- Three-year recidivism probability.

Temporal differentiation allows judicial professionals to distinguish between short- and medium-term projections.

For example:

- A moderate short-term probability combined with elevated three-year probability may indicate delayed structural risk factors.
- A high short-term probability may reflect immediate reintegration instability.

Judges should consider how temporal framing aligns with sentencing and supervision decisions.

The Portuguese model underwent repeated validation using stratified data splits. Performance and fairness metrics were evaluated jointly to ensure balanced optimisation. Although predictive performance varies across subgroups and offence types, the model maintains stable classification behaviour within defined thresholds.

ANNEX A: SYSTEM USE CASES

GREEK MODEL CASE

Greek Use Case 1

Variable	Value
Gender	Female
Nationality	Greek
Education level	Higher Education
Family status	Married
Employment	Unemployed
Age of release	38
Penal situation	Imprisonment
Days in prison	502
Sentence length	1095
Crime category	Acts against property involving violence or threat against a person

System Output

Metric	Value
Risk classification (general)	Medium
Risk classification (3 years)	Medium
Probability (general)	68%
Probability (3 years)	67%

Explainability Summary

The prediction reflects a balanced interaction between institutional and socio-demographic variables. The duration of incarceration (days in prison) contributes to reducing the predicted risk, indicating a stabilizing effect associated with longer exposure to structured correctional environments. In contrast, sentence length acts as a strong risk-increasing factor, suggesting that more severe judicial outcomes are associated with higher likelihoods of recidivism.

Age at release contributes to lowering the risk, reflecting the well-established trend of decreasing recidivism with age. Educational attainment, despite being at a higher level, introduces a slight increase in risk within the model’s learned patterns, while other socio-demographic variables, such as marital and employment status, play a secondary role. Overall, the model produces a medium-risk classification due to the counterbalancing nature of institutional severity and stabilizing temporal factors.

AI-Powered Recidivism Prediction Tool

Risk Assessment

★ RESULTS

1. Risk of reincarceration (classification)

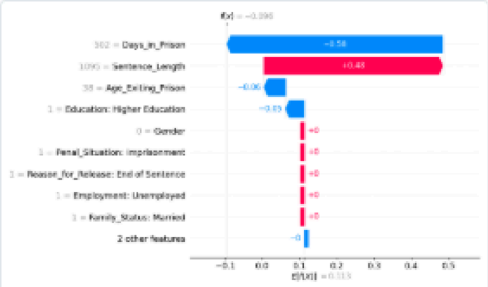
MEDIUM risk of recidivism	**MEDIUM** risk of recidivism within 3 years

2. Probability

68% risk of recidivism	**67%** risk of recidivism within 3 years

3. Explanation

DIAGRAM



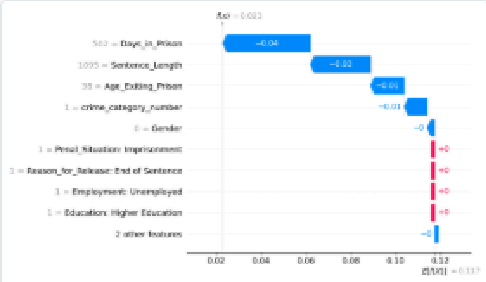
TEXT

The most significant factors influencing this prediction are: Days_in_Prison (decreases risk), Sentence_Length (increases risk), Age_Exiting_Prison (decreases risk).

- **Days_in_Prison** — Decrease Risk (SHAP -0.58)
- **Sentence_Length** — Increase Risk (SHAP 0.48)
- **Age_Exiting_Prison** — Decrease Risk (SHAP -0.06)
- **Education: Higher Education** — Decrease Risk (SHAP -0.05)

recidivism

DIAGRAM



TEXT

The most significant factors influencing this prediction are: Days_in_Prison (decreases risk), Sentence_Length (decreases risk), Age_Exiting_Prison (decreases risk).

- **Days_in_Prison** — Decrease Risk (SHAP -0.04)
- **Sentence_Length** — Decrease Risk (SHAP -0.03)
- **Age_Exiting_Prison** — Decrease Risk (SHAP -0.01)
- **crime_category_number** — Decrease Risk (SHAP -0.01)

within 3 years

Input fields	Input values
Gender	Female
Nationality	Greek
Education level	Higher Education
Family status	Married
Employment	Unemployed
Age of release	38
Penal situation	Imprisonment
Days in prison	502
Sentence length	1095
Crime category	Acts against property involving violence or threat against a person

Export Input and Results

Greek Use Case 2

Variable	Value
Gender	Male
Nationality	Foreigner
Education level	Lower Secondary School
Family status	Single
Employment	Unemployed
Age of release	38
Penal situation	Conviction
Days in prison	139
Sentence length	3650
Crime category	Acts against property involving violence or threat against a person

System Output

Metric	Value
Risk classification (general)	Medium
Risk classification (3 years)	Medium
Probability (general)	73%
Probability (3 years)	48%

Explainability Summary

In this case, the prediction is primarily driven by sentence-related variables. Sentence length emerges as the dominant risk-increasing factor, significantly influencing the model's assessment due to its association with more severe criminal behavior patterns. This effect outweighs the mitigating influence of other variables.

At the same time, time spent in prison contributes to reducing the predicted risk, indicating a partial stabilizing effect. Age at release also slightly decreases the risk, consistent with broader recidivism trends. Nationality and educational background introduce minor adjustments within the model's internal weighting but do not significantly alter the overall outcome.

The result is a medium-risk classification, where the strong upward pressure from sentence severity is partially offset by temporal and demographic stabilizing factors. The divergence between general and 3-year probability reflects differences in short-term versus longer-term predictive behavior of the model.

AI-Powered Recidivism Prediction Tool

Risk Assessment

★ RESULTS

1. Risk of reincarceration (classification)

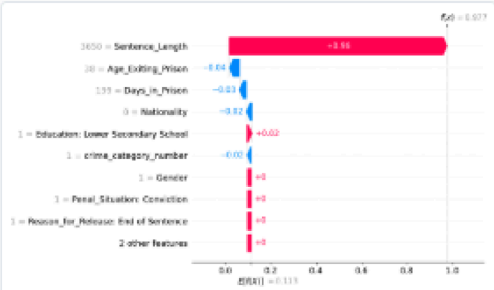
MEDIUM risk of recidivism	**MEDIUM** risk of recidivism within 3 years

2. Probability

73% risk of recidivism	**48%** risk of recidivism within 3 years

3. Explanation

DIAGRAM



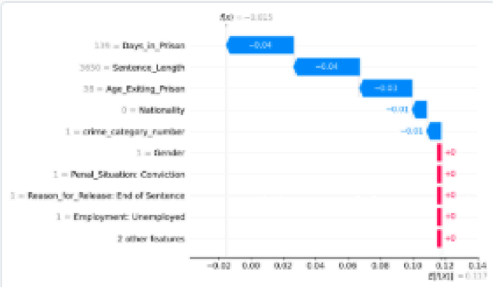
TEXT

The most significant factors influencing this prediction are: Sentence_Length (increases risk), Age_Exiting_Prison (decreases risk), Days_in_Prison (decreases risk).

- **Sentence_Length** — Increase Risk (SHAP 0.96)
- **Age_Exiting_Prison** — Decrease Risk (SHAP -0.04)
- **Days_in_Prison** — Decrease Risk (SHAP -0.03)
- **Nationality** — Decrease Risk (SHAP -0.02)

recidivism

DIAGRAM



TEXT

The most significant factors influencing this prediction are: Days_in_Prison (decreases risk), Sentence_Length (decreases risk), Age_Exiting_Prison (decreases risk).

- **Days_in_Prison** — Decrease Risk (SHAP -0.04)
- **Sentence_Length** — Decrease Risk (SHAP -0.04)
- **Age_Exiting_Prison** — Decrease Risk (SHAP -0.03)
- **Nationality** — Decrease Risk (SHAP -0.01)

within 3 years

Input fields	Input values
Gender	Male
Nationality	Foreigner
Education level	Lower Secondary School
Family status	Single
Employment	Unemployed
Age of release	38
Penal situation	Conviction
Days in prison	139
Sentence length	3650
Crime category	Acts against property involving violence or threat against a person

Export Input and Results

BULGARIAN MODEL CASE

Bulgarian Use Case 1

Variable	Value
Age at entering prison	33
Age at exiting prison	32
Gender	Female
Nationality	Bulgarian
Level of education	Secondary
Marital status	Divorced
Siblings/family	No
Prison status	measure-detained
Days in prison	823
Days of penalty	240
Sentence for multiple crimes	No
Exemption from serving the sentence	Yes
Revocation of the release	No
Sentence fulfilled	Yes
Interruption of implementation	No
Conditional early release	No
Reducing punishment with work	No

System Ouput

Metric	Value
Risk classification	High
Probability	63%

Explainability Summary

The prediction is primarily driven by sentence completion, which appears as the strongest risk-increasing factor within the model. This suggests that, in the Bulgarian dataset, completed sentences—especially under specific institutional pathways—are associated with patterns of re-entry into the correctional system. Additional upward pressure is introduced by penalty-related variables, such as days of penalty, as well as revocation-related indicators, reflecting procedural and judicial complexity.

At the same time, several variables contribute to reducing the predicted risk. Exemption from serving the sentence and reducing punishment with work both act as stabilizing factors, indicating forms of institutional mitigation or alternative enforcement pathways. Educational level also contributes to lowering the risk, while the absence of family support introduces only a minor effect.

Overall, the model produces a high-risk classification due to the dominance of institutional and sentence-related escalation factors over mitigating elements.

★ RESULTS

1. Risk of reincarceration (classification)

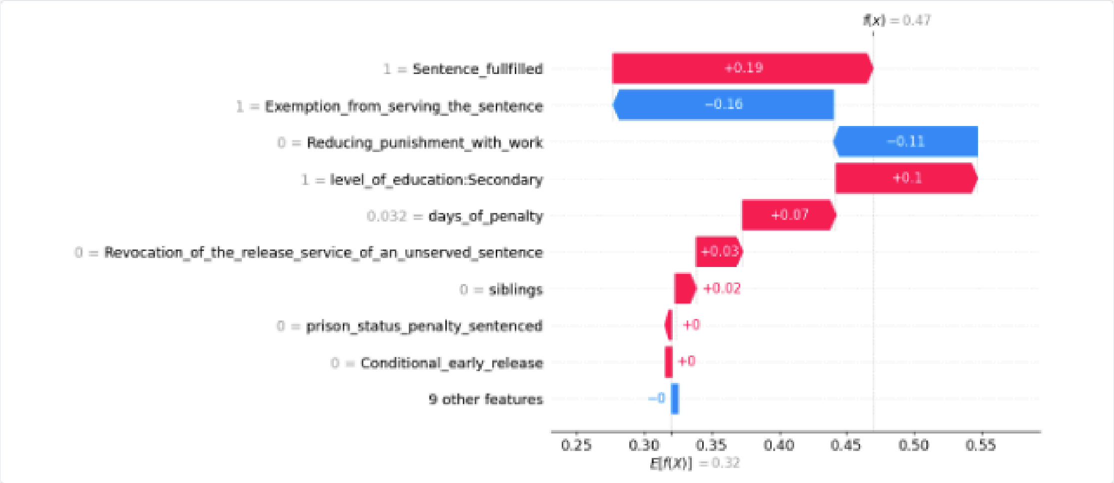
HIGH
risk of reincarceration

2. Probability

63%
risk of reincarceration

3. Explanation

DIAGRAM



TEXT

The most significant factors influencing this prediction are: Sentence_fullfilled (increases risk), Exemption_from_serving_the_sentence (decreases risk), Reducing_punishment_with_work (decreases risk).

- **Sentence_fullfilled** — Increase Risk (SHAP 0.19)
- **Exemption_from_serving_the_sentence** — Decrease Risk (SHAP -0.16)
- **Reducing_punishment_with_work** — Decrease Risk (SHAP -0.11)
- **level_of_education:Secondary** — Increase Risk (SHAP 0.10)
- **days_of_penalty** — Increase Risk (SHAP 0.07)
- **Revocation_of_the_release_service_of_an_unserved_sentence** — Increase Risk (SHAP 0.03)

Input fields	Input values
Age at entering prison	33
Age at exiting prison	32
BG — Gender	Female
BG — Nationality	Bulgarian
Level of education	Secondary
Marital status	Divorced
Siblings/family	No
Prison status	measure-detained
Days in prison	323
Days of penalty	240
Sentence for multiple crimes	No
Exemption from serving the sentence	Yes
Revocation of the release	No
Sentence fulfilled	Yes
Interruption of implementation	No
Conditional early release	No
Reducing punishment with work	No

Bulgarian Use Case 2

Variable	Value
Age at entering prison	37
Age at exiting prison	40
Gender	Male
Nationality	Bulgarian
Level of education	Primary
Marital status	Divorced
Siblings/family	Yes
Prison status	penalty-sentenced
Days in prison	926
Days of penalty	1275
Sentence for multiple crimes	No
Exemption from serving the sentence	Yes
Revocation of the release	Yes
Sentence fulfilled	Yes
Interruption of implementation	No
Conditional early release	Yes
Reducing punishment with work	No

System Output

Metric	Value
Risk classification	Low
Probability	17%

Explainability Summary

In this case, the model produces a low-risk classification, driven primarily by stabilizing institutional and behavioral indicators. While sentence fulfillment appears as a risk-increasing factor, its influence is outweighed by several mitigating variables. The presence of structured release mechanisms, including conditional early release and exemption from serving the sentence, contributes to reducing the predicted risk.

Additional downward pressure is introduced by nationality and educational level, as well as the absence of multiple criminal charges. The extended time spent in prison also contributes to stabilisation, suggesting a reduced likelihood of immediate re-entry. Family presence plays a secondary but supportive role in lowering the risk.

Overall, the model identifies this case as low risk due to the combined effect of institutional mitigation measures and the absence of compounding risk factors, demonstrating how the Bulgarian model prioritizes procedural and correctional pathway indicators in its assessment.

★ RESULTS

1. Risk of reincarceration (classification)

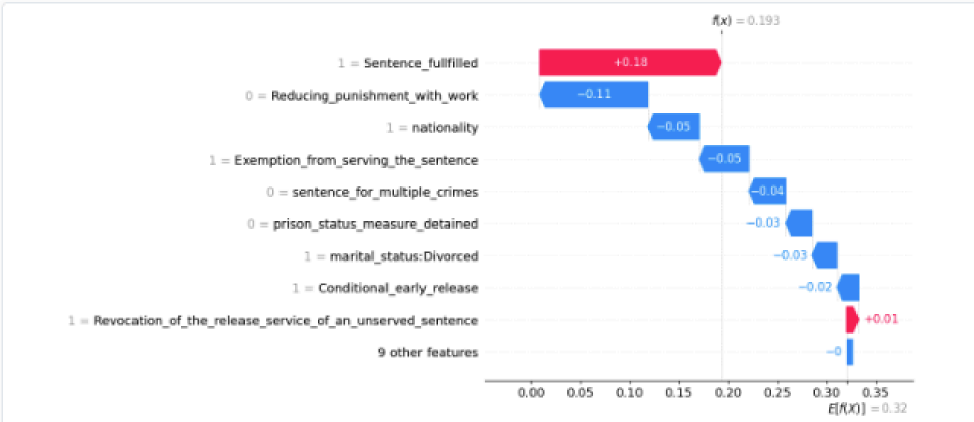
LOW
risk of reincarceration

2. Probability

17%
risk of reincarceration

3. Explanation

DIAGRAM



TEXT

The most significant factors influencing this prediction are: Sentence_fullfilled (increases risk), Reducing_punishment_with_work (decreases risk), nationality (decreases risk).

- **Sentence_fullfilled** — Increase Risk (SHAP 0.18)
- **Reducing_punishment_with_work** — Decrease Risk (SHAP -0.11)
- **nationality** — Decrease Risk (SHAP -0.05)
- **Exemption_from_serving_the_sentence** — Decrease Risk (SHAP -0.05)
- **sentence_for_multiple_crimes** — Decrease Risk (SHAP -0.04)
- **prison_status_measure_detained** — Decrease Risk (SHAP -0.03)

Input fields	Input values
Age at entering prison	37
Age at exiting prison	40
BG — Gender	Male
BG — Nationality	Bulgarian
Level of education	Primary
Marital status	Divorced
Siblings/family	Yes
Prison status	penalty-sentenced
Days in prison	926
Days of penalty	1275
Sentence for multiple crimes	No
Exemption from serving the sentence	Yes
Revocation of the release	Yes
Sentence fulfilled	Yes
Interruption of implementation	No
Conditional early release	Yes
Reducing punishment with work	No

Portuguese Model Case

Portuguese Use Case 1

Variable	Value
Gender	Female
Age of release	40–49
Education level	3rd Basic (7–9 years)
Nationality	Portuguese
Sentence length	15 to 25 years
Crime category	Crimes against property
Type of crime	Simple and aggravated fraud/theft
Management complexity	High
Reason for release	Conditional release

System Output

Metric	Value
Risk classification (general)	Low
Risk classification (3 years)	Low
Probability (general)	19%
Probability (3 years)	6%

Explainability Summary

The prediction is primarily driven by age-related factors, with the age group acting as the strongest risk-reducing variable, reflecting lower recidivism likelihood in more mature populations. Type of crime also contributes to reducing the predicted risk, indicating that the specific nature of the offense is associated with lower recurrence patterns within the Portuguese dataset.

At the same time, the broader crime category introduces a slight increase in risk, suggesting that property-related offenses retain some predictive weight. The conditional release mechanism contributes marginally to increasing the risk, reflecting potential uncertainty in post-release supervision contexts. Management complexity contributes to reducing risk in the longer-term prediction, indicating that structured case management may have a stabilizing effect. Overall, the model produces a low-risk classification due to the dominance of age-related and offense-type stabilizing factors.

AI-Powered Recidivism Prediction Tool

Risk Assessment

★ RESULTS

1. Risk of reincarceration (classification)

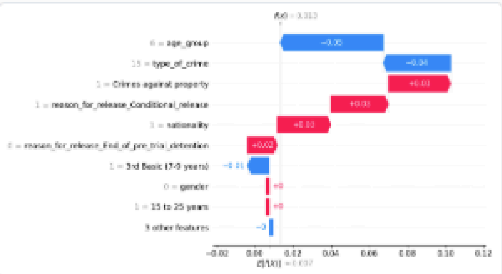
LOW	LOW
risk of recidivism	risk of recidivism within 3 years

2. Probability

19%	6%
risk of recidivism	risk of recidivism within 3 years

3. Explanation

DIAGRAM



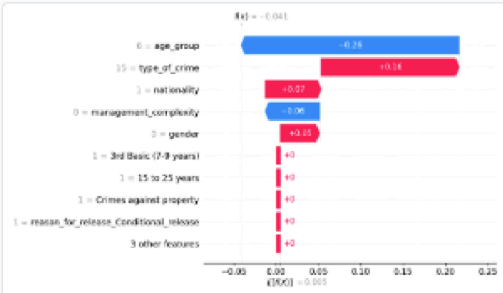
TEXT

The most significant factors influencing this prediction are: age_group (decreases risk), type_of_crime (decreases risk), Crimes against property (increases risk).

- **age_group** — Decrease Risk (SHAP -0.05)
- **type_of_crime** — Decrease Risk (SHAP -0.04)
- **Crimes against property** — Increase Risk (SHAP 0.03)
- **reason_for_release_Conditional_release** — Increase Risk (SHAP 0.03)

recidivism

DIAGRAM



TEXT

The most significant factors influencing this prediction are: age_group (decreases risk), type_of_crime (increases risk), nationality (increases risk).

- **age_group** — Decrease Risk (SHAP -0.26)
- **type_of_crime** — Increase Risk (SHAP 0.16)
- **nationality** — Increase Risk (SHAP 0.07)
- **management_complexity** — Decrease Risk (SHAP -0.06)

within 3 years

Input fields	Input values
PT — Gender	Female
PT — Age of release	40, 49
PT — Education level	3rd Basic (7-9 years)
PT — Nationality	Portuguese
PT — Sentence length	15 to 25 years
PT — Crime category	Crimes against property
PT — Type of crime	Simple and aggravated fraud\theft
PT — Management complexity	High
PT — Reason for release	Conditional release

Export Input and Results

Portuguese Use Case 2

Variable	Value
Gender	Male
Age of release	30–39
Education level	3rd Basic (7–9 years)
Nationality	Portuguese
Sentence length	3 to 6 years
Crime category	Crimes against persons
Type of crime	Assault
Management complexity	High
Reason for release	Other reasons

System Output

Metric	Value
Risk classification (general)	High
Risk classification (3 years)	Low
Probability (general)	70%
Probability (3 years)	24%

Explainability Summary

This case demonstrates a differentiated temporal prediction, where short-term risk is significantly higher than longer-term risk. The type of crime acts as a strong risk-reducing factor, while the age group contributes to lowering the predicted risk, indicating that the individual's demographic profile partially offsets risk exposure.

However, nationality and release-related conditions introduce upward pressure in the prediction, particularly in the short-term assessment. The reason for release contributes to increasing the predicted risk, suggesting that non-standard release pathways may be associated with higher uncertainty. In the 3-year horizon, the model shifts its weighting, with crime-related and age-related factors becoming dominant in reducing risk, leading to a lower classification.

Overall, the model highlights how temporal dynamics affect risk estimation, demonstrating that short-term and long-term recidivism patterns may differ significantly depending on the interaction of demographic, judicial, and offense-related variables.

Risk Assessment

★ RESULTS

1. Risk of reincarceration (classification)

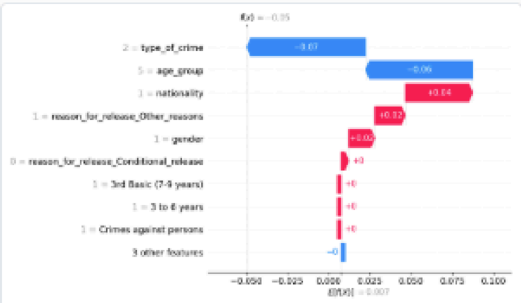
HIGH risk of recidivism	**LOW** risk of recidivism within 3 years

2. Probability

70% risk of recidivism	**24%** risk of recidivism within 3 years

3. Explanation

DIAGRAM



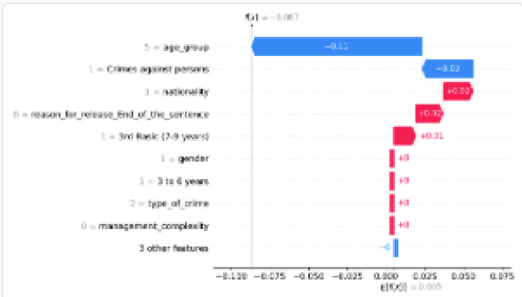
TEXT

The most significant factors influencing this prediction are: type_of_crime (decreases risk), age_group (decreases risk), nationality (increases risk).

- **type_of_crime** — Decrease Risk (SHAP -0.07)
- **age_group** — Decrease Risk (SHAP -0.06)
- **nationality** — Increase Risk (SHAP 0.04)
- **reason_for_release_Other_reasons** — Increase Risk (SHAP 0.02)

recidivism

DIAGRAM



TEXT

The most significant factors influencing this prediction are: age_group (decreases risk), Crimes against persons (decreases risk), nationality (increases risk).

- **age_group** — Decrease Risk (SHAP -0.11)
- **Crimes against persons** — Decrease Risk (SHAP -0.03)
- **nationality** — Increase Risk (SHAP 0.02)
- **reason_for_release_End_of_the_sentence** — Increase Risk (SHAP 0.02)

within 3 years

Input fields	Input values
PT — Gender	Male
PT — Age of release	30, 39
PT — Education level	3rd Basic (7-9 years)
PT — Nationality	Portuguese
PT — Sentence length	3 to 6 years
PT — Crime category	Crimes against persons
PT — Type of crime	Assault
PT — Management complexity	High
PT — Reason for release	Other reasons

Export Input and Results