University of Essex Online
MSc Artificial Intelligence

Mitigating machine learning bias in criminal justice: An ontological approach to predicting recidivism in England and Wales

Author: Leigh Feaviour

First Supervisor: Godfried Williams

Second Supervisor: Samuel Danso

September 2024

Agenda

- Abstract
- Research Question
- Professional and Ethical Considerations
- Literature Review
- Methodology
- National Institute of Justice (NIJ) Statistical Analysis
- NIJ Ontology Design with Feature Correlations
- NIJ Ontology Interrogation
- Ministry of Justice (MoJ) Ontology Design
- MoJ Ontology Interrogation
- Evaluation
- Conclusions and Recommendations
- References

Abstract

Recidivism is when someone who has been convicted of a crime reoffends.

Machine learning is used to predict recidivism, but with examples of racial bias such as Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) in America.

A thorough literature review found complex relationships between recidivism prediction models and sources of bias, including the data, feature selection and the chosen performance metrics.

Ontology is proven to be an effective mitigation to biases by storing metaknowledge about the correlation of features to protected characteristics so data scientists can select features with greater correlation to recidivism and lower correlation to race (transparency of input) and ensuring that results conform to expected distributions (transparency of output).

Research Question

“Can ontology mitigate bias when using machine learning to predict recidivism?”

Aim:

Machine learning is used to predict recidivism, but previous studies have indicated ethical issues such as racial bias. This study will show if biases can be identified and mitigated with the use of ontology by creating an ontology of criminal justice in England and Wales. Features will be identified for safely predicting recidivism and features to be used with caution. Furthermore, the protected characteristic profiles can be compared with predicted profiles to check for parity.

Objectives:

- Identify features that predict recidivism.
- Identify features that potentially introduce biases when predicting recidivism.
- Assess if recidivism varies between characteristics such as ethnicity, gender and age.
- Create an ontology of criminal justice using available metadata and illustrate how the ontology can manage the features to reduce biases, as well as highlight potential biases in the output to further mitigate bias risks.

Professional and Ethical Considerations

	Code of conduct. You shall:	Compliance statement for this project
Public Interest	have due regard for public health, privacy, security and wellbeing of others and the environment.	N/A. No opportunity
	have due regard for the legitimate rights of Third Parties.	Compliant. No direct third-party interaction, but rights always considered
	conduct your professional activities without discrimination on the grounds of sex, sexual orientation, marital status, nationality, colour, race, ethnic origin, religion, age or disability, or of any other condition or requirement.	Compliant. This project aims to reduce discrimination
	promote equal access to the benefits of IT and seek to promote the inclusion of all sectors in society wherever opportunities arise.	N/A. The project does not discuss who would use the solution
Professional Competence and Integrity	only undertake to do work or provide a service that is within your professional competence.	Compliant. The project is an extension of learning undertaken on the MSc
	NOT claim any level of competence that you do not possess.	Compliant
	develop your professional knowledge, skills and competence on a continuing basis, maintaining awareness of technological developments, procedures, and standards that are relevant to your field.	Compliant. The project, including this report, are examples
	ensure that you have the knowledge and understanding of Legislation and that you comply with such Legislation, in carrying out your professional responsibilities.	Compliant. Legislation has been considered and discussed
	respect and value alternative viewpoints and, seek, accept and offer honest criticisms of work.	Compliant. Input from supervisors has shaped the scope
	avoid injuring others, their property, reputation, or employment by false or malicious or negligent action or inaction.	Compliant
	reject and will not make any offer of bribery or unethical inducement.	Compliant
Duty to Relevant Authority	carry out your professional responsibilities with due care and diligence in accordance with the Relevant Authority's requirements whilst exercising your professional judgement at all times.	Compliant
	seek to avoid any situation that may give rise to a conflict of interest between you and your Relevant Authority.	Compliant
	accept professional responsibility for your work and for the work of colleagues who are defined in a given context as working under your supervision.	Compliant
	NOT disclose or authorise to be disclosed, or use for personal gain, or to benefit a third party, confidential information except with the permission of your Relevant Authority, or as required by Legislation.	Compliant. No personal data used in the project. All data was public domain.
	NOT misrepresent or withhold information on the performance of products, systems or services (unless lawfully bound by a duty of confidentiality not to disclose such information), or take advantage of the lack of relevant knowledge or inexperience of others.	Compliant
Duty to the Profession	accept your personal duty to uphold the reputation of the profession and not take any action which could bring the profession into disrepute.	Compliant. All project activities were professional
	seek to improve professional standards through participation in their development, use and enforcement.	N/A. No opportunity
	uphold the reputation and good standing of BCS, the Chartered Institute for IT.	Compliant. The project addresses bias which is reputationally positive
	act with integrity and respect in your professional relationships with all members of BCS and with members of other professions with whom you work in a professional capacity.	Compliant. All research and supervisor discussions were respectful and professional
	encourage and support fellow members in their professional development.	N/A. No opportunity

Ethical and professional considerations were assessed using the BCS code of conduct (BCS, 2022)

Literature Review

Recidivism prediction

- America: for parole decisions using age, intelligence, nationality and criminal history since 1920s (Borden, 1928)
- America: Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) widely used (Equivant, 2019) but shows bias against black people (Angwin et al., 2016)
- Canada: Statistical Information on Recidivism – Revised (SIR-R1) using 15 features (Nafekh & Motiuk, 2002)
- England & Wales: Offender Group Reconviction Scale (OGRS) actuarial tool (HM Prison & Probation Service, 2023)
- Various machine learning solutions tested (Curtis, 2018; Kovalchuk et al., 2023; Lin et al., 2020; Tollenaar & van der Heijden, 2013; Wang et al., 2010; Zeng et al., 2017)
- Dynamic factors predict recidivism (Andrews & Bonda, 2024; Farrington & West, 1995; Farrington et al., 2017; Osborn, 1980) but are rarely used

Sources of bias

- Performance metrics chosen (Caton & Haas, 2020)
- Feature selection (Angwin et al., 2016)
- Data (Biddle, 2022)

Explainability

- Transparency is important for ethical machine learning models (Walmsley, 2021)
- Better to use an explainable model that try to explain a black-box model (Rudin, 2019)

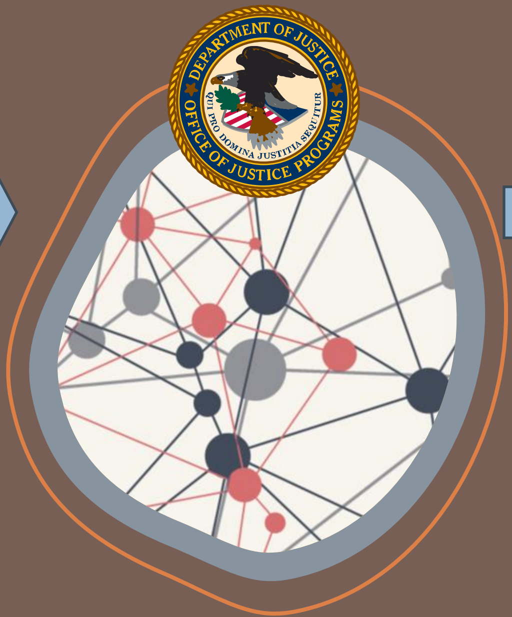
Challenges

- Correcting ethical imbalances decreases accuracy as ethical compliance increases (Squadrone et al., 2022)
- It is rarely possible to calibrate within groups, balance the positive class, and balance the negative class simultaneously (Kleinberg et al., 2016)
- Age is a good predictor (Bushway & Piehl, 2007; Kleiman et al., 2007; Stevenson & Slobogin, 2018), but it is static and cannot be influenced
- Men and women have different recidivism rates so differentiating increases accuracy and fairness (Skeem & Lowenkamp, 2020), but gender is a protected static characteristic

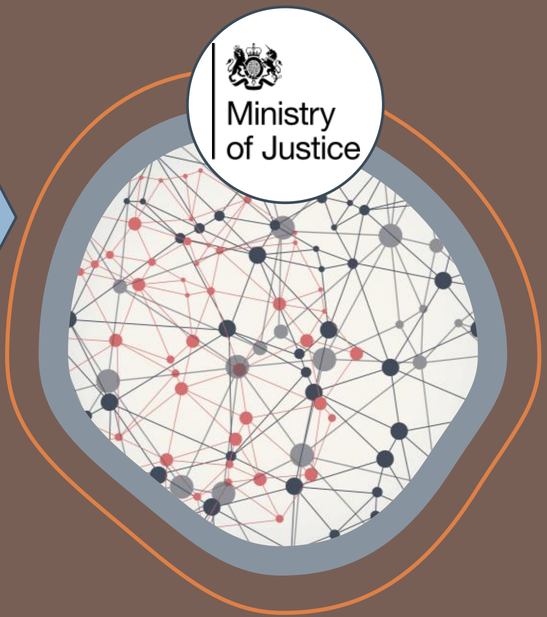
Methodology

NIJ EDA

NIJ Statistical Analysis



NIJ Ontology



MoJ Ontology

Tool:	Python	R	Protégé	Protégé
Outcome:	Form hypotheses	Test hypotheses	Simple proof of concept	Real-world proof of concept
Granularity:	High level	Detailed analysis	Basic ontology, real data	Complex ontology, dummy data

NIJ Dataset (25,835 records) from National Institute of Justice (N.D.)

Metadata from Crown Courts, Magistrate Courts, Prison Service and Probation Service in England and Wales from Ministry of Justice (2020)

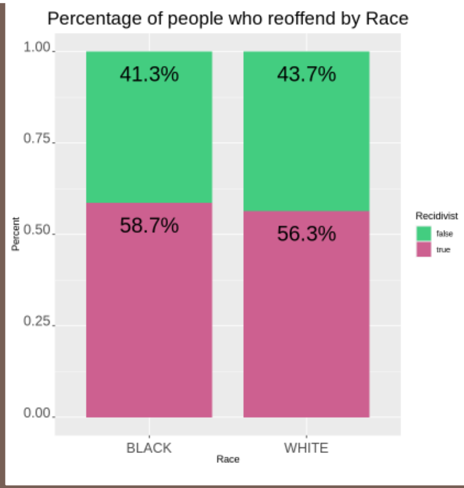
NIJ Statistical Analysis

Hypotheses:

1. There is a difference in recidivism by race, α=0.01 — Null hypothesis rejected
2. There is a difference in recidivism by gender, α=0.01 — Null hypothesis rejected
3. There is a difference in recidivism by age, α=0.01 — Null hypothesis rejected
4. There is a difference in offender age by race, α=0.01 — Null hypothesis rejected

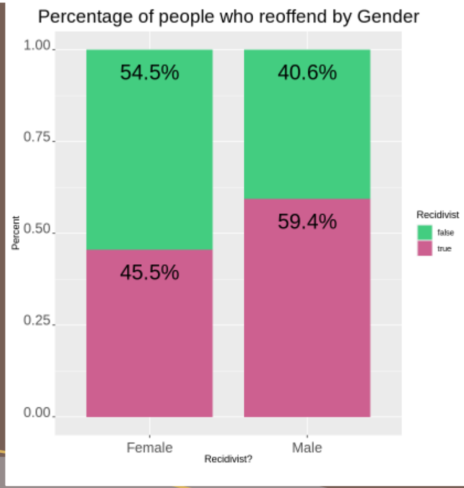
```
       BLACK WHITE
false  6134  4797
true   8713  6191
        Pearson's Chi-squared test with Yates' continuity correction

data:  table(NIJ_orig$Recidivism_Within_3years, NIJ_orig$Race)
X-squared = 14.094, df = 1, p-value = 0.0001739
```

1



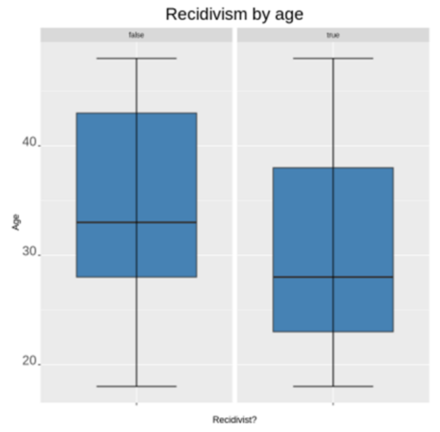

```
        Male Female
false   9206   1725
true   13462   1442
        Pearson's Chi-squared test with Yates' continuity correction

data:  table(NIJ_orig$Recidivism_Within_3years, NIJ_orig$Gender)
X-squared = 217.99, df = 1, p-value < 2.2e-16
```

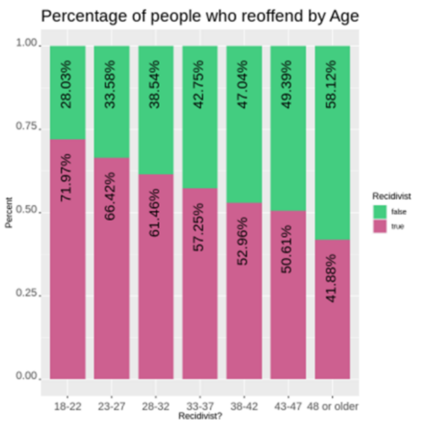
2

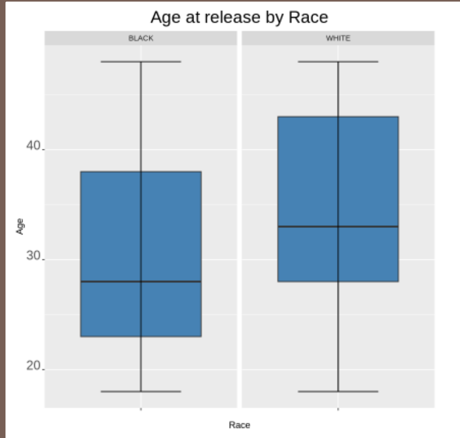



```
        Wilcoxon rank sum test with continuity correction

data:  Age_at_Release by Recidivism_Within_3years
W = 98018098, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
        Welch Two Sample t-test

data:  Age_at_Release by as_factor(Recidivism_Within_3years)
t = 28.822, df = 22988, p-value < 2.2e-16
alternative hypothesis: true difference in means between group false and group true is not equal to 0
95 percent confidence interval:
 3.145913 3.605023
sample estimates:
mean in group false  mean in group true
           34.53737            31.16190
```

4





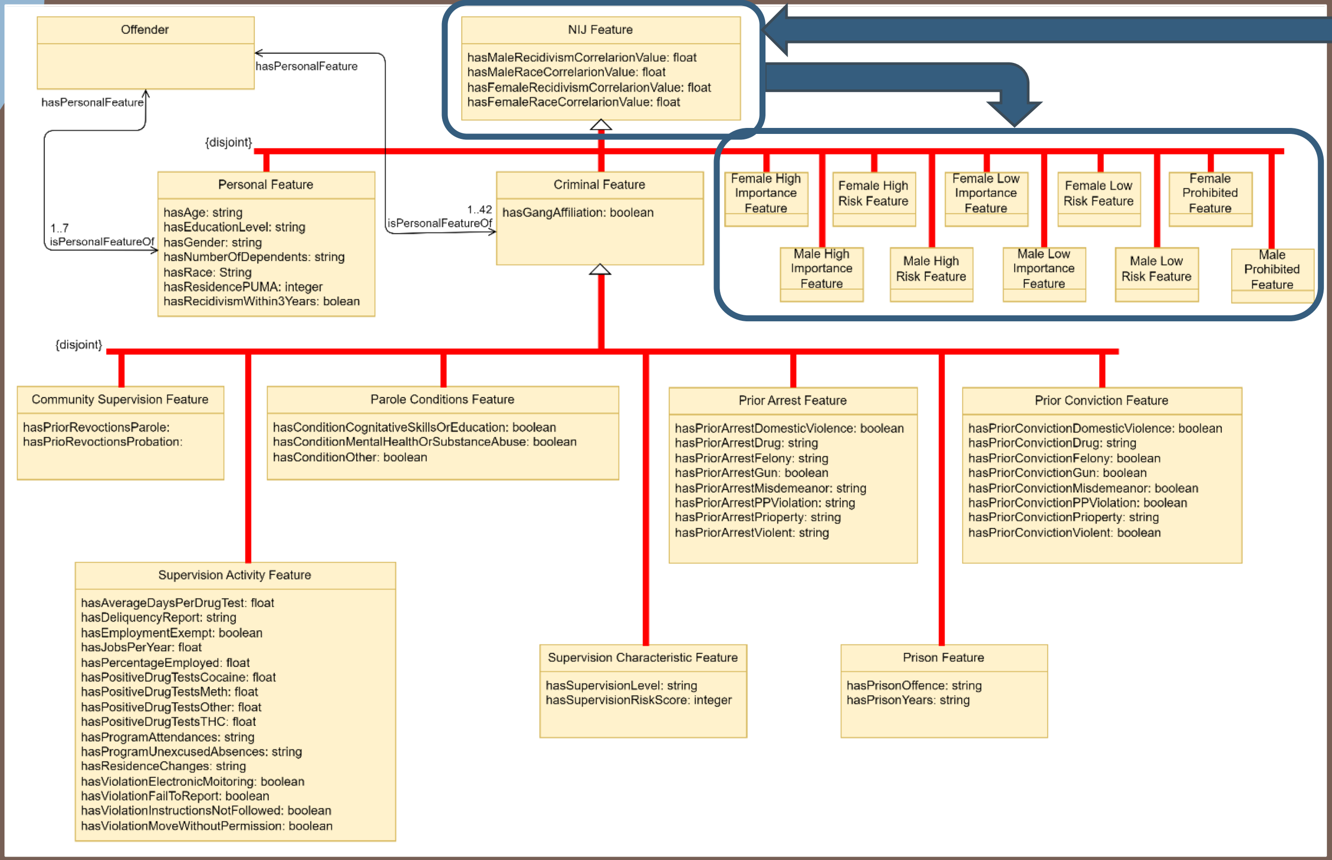


3

```
        Wilcoxon rank sum test with continuity correction

data:  Age_at_Release by Race
W = 69826119, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
        Welch Two Sample t-test

data:  Age_at_Release by as_factor(Race)
t = -19.514, df = 24010, p-value < 2.2e-16
alternative hypothesis: true difference in means between group BLACK and group WHITE is not equal to 0
95 percent confidence interval:
 -2.509302 -2.051234
sample estimates:
mean in group BLACK mean in group WHITE
           31.62026            33.90053
```

Every features was tested for correlation with recidivism and race by gender, Spearman's Rho, α=0.01 (Kim & Choi, 2021)
Age correlates with recidivism (Bushway & Piehl, 2007; Kleiman et al., 2007; Stevenson & Slobogin, 2018) and had r_s -0.177 with recidivism with p 2.2e-16 so r_s 0.1; weak correlation (Xiao et al., 2016) was selected as the cut-off for high/low risk/importance

NIJ Ontology Design with Feature Correlations



Feature	Correlation with: Everyone	Male		Female	
	Recidivism	Recidivism	Race	Recidivism	Race
Age_at_Release	0.176	0.177	0.121	0.133	0.072
Residence_PUMA	0.025	0.026	0.139	0	0.187
Gang_Affiliated	0.185	0.185	0.086	N/A	N/A
Supervision_Risk_Score_First	0.180	0.185	0.053	0.146	0.046
Supervision_Level_First	0.061	0.053	0	0.069	0
Education_Level	0.088	0.088	0.057	0	0
Dependents	0.031	0.031	0.096	0	0.064
Prison_Offense	0.018	0.024	0.033	0	0.260
Prison_Years	0.130	0.134	0.066	0.186	0.109
Prior_Arrest_Episodes_Felony	0.199	0.187	0.025	0.262	0
Prior_Arrest_Episodes_Misd	0.178	0.161	0.094	0.279	0
Prior_Arrest_Episodes_Violent	0.065	0.055	0.111	0	0.213
Prior_Arrest_Episodes_Property	0.182	0.181	0.103	0.233	0
Prior_Arrest_Episodes_Drug	0.081	0.071	0	0.107	0.279
Prior_Arrest_Episodes_PPViolationCharges	0.229	0.218	0.063	0.303	0.067
Prior_Arrest_Episodes_DVCharges	0.066	0.062	0.052	0.052	0
Prior_Arrest_Episodes_GunCharges	0.044	0.036	0.104	0	0
Prior_Conviction_Episodes_Felony	0.105	0.094	0.032	0.169	0.047
Prior_Conviction_Episodes_Misd	0.175	0.160	0.070	0.247	0
Prior_Conviction_Episodes_Viol	0.047	0.043	0.088	0	0.161
Prior_Conviction_Episodes_Prop	0.161	0.157	0.104	0.232	0.073
Prior_Conviction_Episodes_Drug	0.065	0.059	0	0.077	0.235
Prior_Conviction_Episodes_PPViolationCharges	0.096	0.088	0.050	0.137	0
Prior_Conviction_Episodes_DomesticViolenceCharges	0.059	0.057	0.017	0	0
Prior_Conviction_Episodes_GunCharges	0.031	0.024	0.058	0	0
Prior_Revocations_Parole	0.058	0.051	0.037	0.060	0
Prior_Revocations_Probation	0.039	0.036	0.065	0.076	0.059
Condition_MH_SA	0.114	0.121	0.131	0.149	0.259
Condition_Cog_Ed	0.038	0.050	0.039	0	0
Condition_Other	0	0	0	0	0.065
Violations_ElectronicMonitoring	0.004	0	0.069	0	0.075
Violations_Instruction	0.064	0.058	0.046	0.087	0
Violations_FailToReport	0.030	0.024	0	0.069	0
Violations_MoveWithoutPermission	0.032	0.029	0	0.057	0
Delinquency_Reports	0.041	0.028	0	0.102	0.068
Program_Attendances	0.060	0.065	0.072	0	0.190
Program_UnexcusedAbsences	0.060	0.050	0.043	0.108	0
Residence_Changes	0.054	0.052	0.047	0.079	0
Avg_Days_per_DrugTest	0.011	0	0.078	0	0.135
DrugTests_THC_Positive	0.082	0.078	0.161	0	0.089
DrugTests_Cocaine_Positive	0.011	0	0.128	0	0.121
DrugTests_Meth_Positive	0.055	0.055	0.279	0.091	0.227
DrugTests_Other_Positive	0.004	0	0.121	0.053	0.126
Percent_Days_Employed	0.217	0.217	0.126	0.227	0.059
Jobs_Per_Year	0.074	0.074	0.120	0.088	0.060
Employment_Exempt	0.050	0.048	0.021	0	0

Legend	
	Recidivism correlation ≥ 0.1
	Race correlation ≥ 0.1
	No statitically significant correlation at α=0.01

- Designed around NIJ dataset (National Institute of Justice, N.D.)
- Recidivism and race correlations added by gender for every feature
- Defined classes created to infer high/low risk/importance features by gender
- NIJ data imported with Cellfie scripts

UML notation for ontologies (Bārzdiņš et al., 2010) adapted to include object properties

NIJ Ontology Interrogation

Snap SPARQL Query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX nij: <http://www.semanticweb.org/leigh/ontologies/2024/4/NIJ#>

#Show features where maleRaceCorrelationValue is greater than maleRecidivismCorrelationValue
SELECT ?feature ?maleRaceCorrelationValue ?maleRecidivismCorrelationValue
WHERE {
 ?feature  nij:hasMaleRaceCorrelationValue ?maleRaceCorrelationValue ;
           nij:hasMaleRecidivismCorrelationValue ?maleRecidivismCorrelationValue .

 FILTER(?maleRaceCorrelationValue > ?maleRecidivismCorrelationValue)
}
ORDER BY DESC(?maleRaceCorrelationValue)
```

Execute

?feature	?maleRaceCorrelationValue	?maleRecidivismCorrelationValue
nij:PositiveDrugTestsMeth	0.279	0.055
nij:PostiveDrugTestsTHC	0.161	0.078
nij:ResidencePUMA	0.139	0.026
nij:ConditionMentalHealthOrSubstanceAbuse	0.131	0.121
nij:PositiveDrugTestsCocaine	0.128	0.0
nij:PositiveDrugTestsOther	0.121	0.0
nij:JobsPerYear	0.12	0.074
nij:PriorArrestViolent	0.111	0.055
nij:PriorArrestGun	0.104	0.036
nij:NumberOfDependents	0.096	0.031
nij:PriorConvictionViolent	0.088	0.043
nij:AverageDaysPerDrugTest	0.078	0.0
nij:ProgramAttendances	0.072	0.065
nij:ViolationElectronicMonitoring	0.069	0.0
nij:PriorRevocationsProbation	0.065	0.036
nij:PriorConvictionGun	0.058	0.024
nij:PrisonOffence	0.033	0.024

17 results

Features where correlation with race is higher than correlation with recidivism
Avoid these features!

DL query:

Query (class expression)

MaleHighImportance and MaleHighRiskFeature

Execute | Add to ontology

Query results

Instances (5 of 5)

- Age
- ConditionMentalHealthOrSubstanceAbuse
- PercentDaysEmployed
- PriorArrestProperty
- PriorConvictionProperty

Features with high importance and high risk
Use with caution!

Snap SPARQL Query:

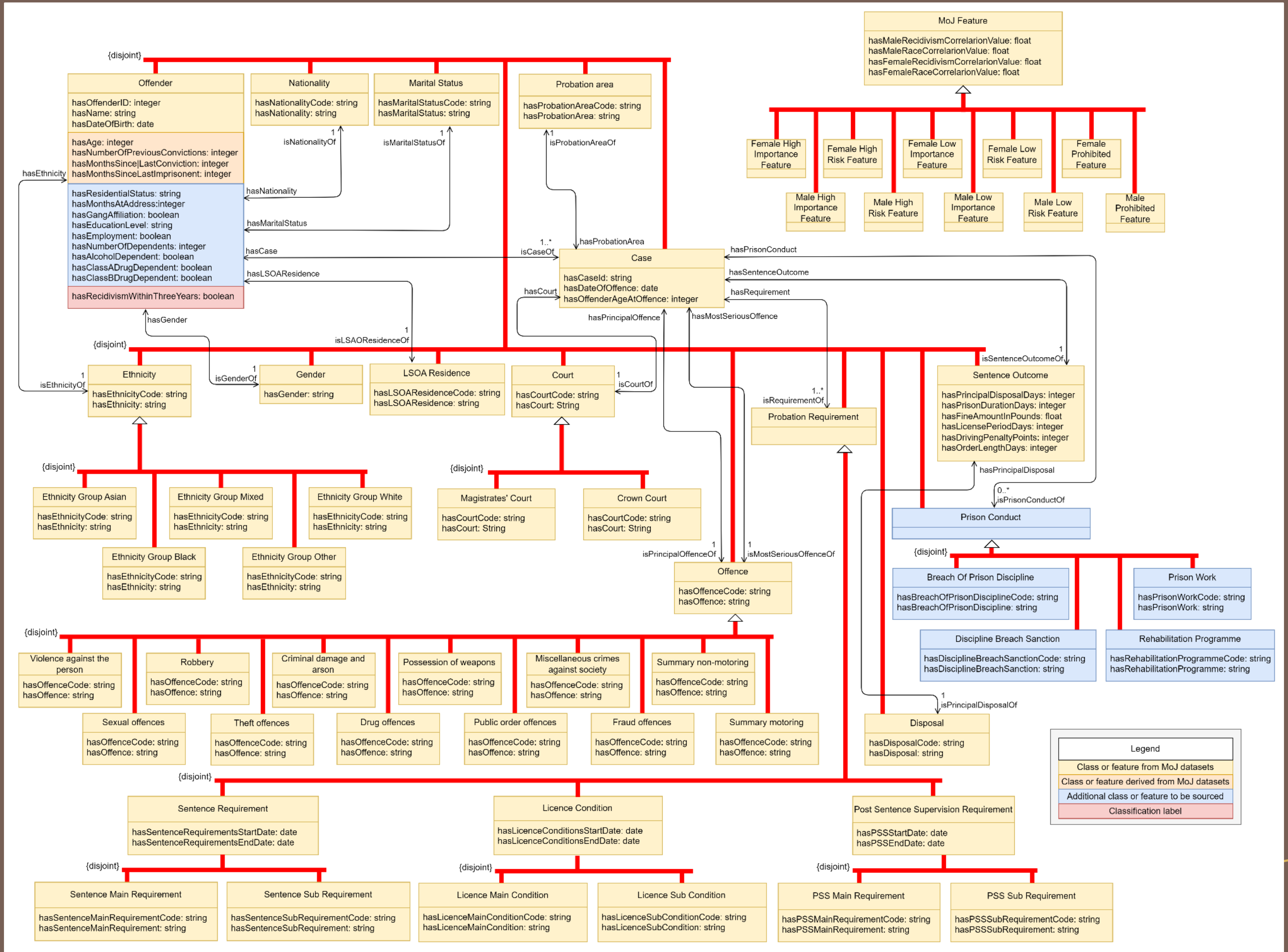
```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX nij: <http://www.semanticweb.org/leigh/ontologies/2024/4/NIJ#>

#Total people and number who reoffended number gender and race
SELECT ?gender ?race (COUNT(?person) AS ?total) (COUNT(?reoffender) AS
?reoffended)
WHERE {
 ?person a nij:Offender ;
          nij:hasGender ?gender ;
          nij:hasRace ?race .

 OPTIONAL {
  ?person nij:hasRecidivismWithin3Years nij:true .
  BIND(IF(BOUND(?person), ?person, "") AS ?reoffender)
 }
}
GROUP BY ?gender ?race
ORDER BY ?gender ?race
```

Execute

?gender	?race	?total	?reoffended
nij:F	nij:BLACK	1082	474
nij:F	nij:WHITE	2085	968
nij:M	nij:BLACK	13765	8239
nij:M	nij:WHITE	8903	5223

Recidivism by race and gender
Machine learning predictions to be validated against actual ratios

MoJ Ontology Design



- Core design based upon metadata from Ministry of Justice (2020)
- Some features need to be calculated outside the ontology e.g. age from DoB
- Additional features added from literature review and NIJ design
- Cellfie scripts from MoJ metadata to import object and data instances
- SWRL rules (Horn clauses) to infer descriptions from codes
- Dummy data created for offenders and cases

- The class holding the correlation properties was separated from the rest of the ontology for significant performance gain (2.5 hours vs. 3 seconds to run reasoner). However, this further separated the ontology knowledge from the correlation metaknowledge

UML notation for ontologies (Bārzdiņš et al., 2010) adapted to include object properties

MoJ Ontology Interrogation

Snap SPARQL Query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX moj: <http://www.semanticweb.org/leigh/ontologies/2024/4/MOJ#>

#Show features where maleRecidivismCorrelationValue is greater than maleRaceCorrelationValue
SELECT ?feature ?maleRecidivismCorrelationValue ?maleRaceCorrelationValue
WHERE {
 ?feature  moj:hasMaleRecidivismCorrelationValue ?maleRecidivismCorrelationValue ;
           moj:hasMaleRaceCorrelationValue ?maleRaceCorrelationValue .

 FILTER(?maleRecidivismCorrelationValue > ?maleRaceCorrelationValue)
}
ORDER BY DESC(?maleRecidivismCorrelationValue)
```

Execute

?feature	?maleRecidivismCorrelationValue	?maleRaceCorrelationValue
moj:MonthsSinceLastImprisonment	0.281	0.142
moj:MonthsSinceLastConviction	0.242	0.132
moj:Employment	0.217	0.126
moj:MonthsAtAddress	0.213	0.095
moj:GangAffiliation	0.185	0.086
moj:Age	0.177	0.121
moj:NumberOfPreviousConvictions	0.164	0.079
moj:OffenderAgeAtOffence	0.162	0.102
moj:PrisonDurationDays	0.162	0.092
moj:RehabilitationProgrammeCode	0.162	0.053
moj:PrincipalOffenceCode	0.161	0.092
moj:MaritalStatusCode	0.153	0.032
moj:PSSSubRequirementCode	0.152	0.056
moj:AlcoholDependent	0.132	0.098
moj:PrisonWorkCode	0.123	0.078
moj:OrderLengthDays	0.112	0.101
moj:LicenceSubConditionCode	0.099	0.045
moj:PSSMainRequirementCode	0.096	0.023
moj:LicenceMainConditionCode	0.092	0.076
moj:PrincipalDisposalDays	0.092	0.021
moj:EducationLevel	0.088	0.057
moj:PrincipalDisposalCode	0.075	0.021
moj:DrivingPenaltyPoints	0.052	0.024
moj:FineAmountPounds	0.043	0.012
moj:ProbationAreaCode	0.02	0.01

25 results

Features where correlation with recidivism is higher than correlation with race
Consider using these features!

Snap SPARQL Query:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX moj: <http://www.semanticweb.org/leigh/ontologies/2024/4/MOJ#>

SELECT ?offender
    (REPLACE(STR(?monthsSinceImprisonmentValue), STR(moj:), "") AS ?monthsSinceImprisonment)
    (REPLACE(STR(?monthsSinceConvictionValue), STR(moj:), "") AS ?monthsSinceConviction)
    (REPLACE(STR(?employmentValue), STR(moj:), "") AS ?employment)
    (REPLACE(STR(?monthsAtAddressValue), STR(moj:), "") AS ?monthsAtAddress)
    (REPLACE(STR(?gangAffiliationValue), STR(moj:), "") AS ?gangAffiliation)
    (REPLACE(STR(?ageValue), STR(moj:), "") AS ?age)
    (REPLACE(STR(?previousConvictionsValue), STR(moj:), "") AS ?previousConvictions)
    (REPLACE(STR(?maritalStatusValue), STR(moj:), "") AS ?maritalStatus)
    (REPLACE(STR(?alcoholDependentValue), STR(moj:), "") AS ?alcoholDependent)
    (REPLACE(STR(?caseValue), STR(moj:), "") AS ?case)
    (REPLACE(STR(?rehabilitationProgrammeValue), STR(moj:), "") AS ?rehabilitationProgramme)
    (REPLACE(STR(?principalOffenceValue), STR(moj:), "") AS ?principalOffence)
    (REPLACE(STR(?PSSSubRequirementValue), STR(moj:), "") AS ?PSSSubRequirement)
    (REPLACE(STR(?prisonWorkValue), STR(moj:), "") AS ?prisonWork)
    (REPLACE(STR(?prisonDurationValue), STR(moj:), "") AS ?prisonDuration)
WHERE {
 ?offender a moj:Offender
        FILTER(REGEX(STR(?offender), CONCAT("^", STR(moj:), "O")))
        OPTIONAL { ?offender moj:hasMonthsSinceLastImprisonment ?monthsSinceImprisonmentValue. }
        OPTIONAL { ?offender moj:hasMonthsSinceLastConviction ?monthsSinceConvictionValue. }
        OPTIONAL { ?offender moj:hasEmployment ?employmentValue. }
        OPTIONAL { ?offender moj:hasMonthsAtAddress ?monthsAtAddressValue. }
        OPTIONAL { ?offender moj:hasGangAffiliation ?gangAffiliationValue. }
        OPTIONAL { ?offender moj:hasAge ?ageValue. }
        OPTIONAL { ?offender moj:hasNumberOfPreviousConvictions ?previousConvictionsValue. }
        OPTIONAL { ?offender moj:hasMaritalStatus ?maritalStatusValue. }
        OPTIONAL { ?offender moj:hasAlcoholDependent ?alcoholDependentValue. }
 ?offender moj:hasCase ?case
        OPTIONAL { ?case moj:hasPrisonDurationDays ?prisonDurationValue. }
        OPTIONAL { ?case moj:hasPrisonDurationDays ?prisonDurationValue. }
        OPTIONAL { ?case moj:hasRehabilitationProgramme ?rehabilitationProgrammeValue. }
        OPTIONAL { ?case moj:hasPrincipalOffence ?principalOffenceValue. }
        OPTIONAL { ?case moj:hasPSSSubRequirement ?PSSSubRequirementValue. }
        OPTIONAL { ?case moj:hasPrisonWork ?prisonWorkValue. }
}
ORDER BY ?offender
```

Execute

?offender	?m...	?m...	?emp...	?m...	?gan...	?age	?p...	?maritalStatus	?alco...	?case	?rehabilitationPr...	?principalOffence	?PSSSubRequ...	?prisonWork	?priso...
moj:O112299926		11	false	13	false	63	4	Divorced or dissolv...	true	moj:C647325366		46 Theft from Shops			
moj:O112299926		11	false	13	false	63	4	Divorced or dissolv...	true	moj:C214350956		46 Theft from Shops			
moj:O112299926		11	false	13	false	63	4	Divorced or dissolv...	true	moj:C691205173		46 Theft from Shops			
moj:O112299926		11	false	13	false	63	4	Divorced or dissolv...	true	moj:C986467303		46 Theft from Shops			
moj:O112299926		11	false	13	false	63	4	Divorced or dissolv...	true	moj:C745987478		46 Theft from Shops			
moj:O141806441		2	false	52	false	59	0	Married or in civil p...	false	moj:C637473147		8.01 Assault occasioning actu...			
moj:O265466553		16	false	15	false	24	0	Married or in civil p...	true	moj:C196164791		46 Theft from Shops			
moj:O341786705	0	8	false	51	false	46	3	Married or in civil p...	true	moj:C527235880		92E.01 Possession of a contr...			
moj:O341786705	0	8	false	51	false	46	3	Married or in civil p...	true	moj:C905424044	The Bridge Progr...	4.6 Causing Death by Careles...			730
moj:O341786705	0	8	false	51	false	46	3	Married or in civil p...	true	moj:C794276181		92D.01 Possession of a contr...			
moj:O341786705	0	8	false	51	false	46	3	Married or in civil p...	true	moj:C830828808		92D.01 Possession of a contr...			
moj:O486623255	0	10	false	13	false	73	3	Widowed	false	moj:C478663808		46 Theft from Shops			
moj:O486623255	0	10	false	13	false	73	3	Widowed	false	moj:C338060576	Building Better R...	8.10 Breach of a restraining or...			730
moj:O486623255	0	10	false	13	false	73	3	Widowed	false	moj:C338060576	Becoming New M...	8.10 Breach of a restraining or...			730
moj:O486623255	0	10	false	13	false	73	3	Widowed	false	moj:C348844983		8.01 Assault occasioning actu...			
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C587659755		46 Theft from Shops			
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C618323430	Living as New Me	34 Robbery	Restorative Ju...	Workshop	365
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C535754077		45 Theft from Vehicle			
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C499266153		46 Theft from Shops			
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C113345537		46 Theft from Shops			
moj:O511985165	299	10	false	36	false	61	5	Single-not married/...	true	moj:C904791425		34 Robbery			
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C375098237		92D.01 Possession of a contr...			
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C851372657		46 Theft from Shops			
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C607333684		92D.01 Possession of a contr...			
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C569872265	Living as New Me	92A.09 Production supply and ...		Servery	1277
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C569872265	Living as New Me	92A.09 Production supply and ...		Cleaner	1277
moj:O547242054	0	1	false	31	false	23	4	Single-not married/...	false	moj:C254596091		92D.01 Possession of a contr...			
moj:O613456052	3	27	false	62	false	69	1	Married or in civil p...	true	moj:C808774974	Identity Matters	53D Fraud by false representa...		Servery	760
moj:O613456052	3	27	false	62	false	69	1	Married or in civil p...	true	moj:C808774974	Identity Matters	53D Fraud by false representa...		Cleaner	760
moj:O613456052	3	27	false	62	false	69	1	Married or in civil p...	true	moj:C723372549		46 Theft from Shops			
moj:O646955585		7	false	9	false	42	0	Single-not married/...	false	moj:C595710701		46 Theft from Shops			
moj:O815709338		4	false	58	false	20	0	Single-not married/...	false	moj:C755833865		46 Theft from Shops			
moj:O815709338		4	false	58	false	20	0	Single-not married/...	false	moj:C178906509		46 Theft from Shops			
moj:O815709338		4	false	58	false	20	0	Single-not married/...	false	moj:C861603678		46 Theft from Shops			
moj:O815709338		4	false	58	false	20	0	Single-not married/...	false	moj:C202280393		46 Theft from Shops			
moj:O815709338		4	false	58	false	20	0	Single-not married/...	false	moj:C112529789		46 Theft from Shops			
moj:O817633970		1	false	37	false	29	4	Single-not married/...	true	moj:C782821873		8.07 Racially or religiously ag...			
moj:O983190851		15	false	3	false	26	0	Married or in civil p...	false	moj:C281645560		8.07 Racially or religiously ag...			

38 results

Export of selected data

Evaluation

Results

EDA and statistical analysis reflected old but valid best practices (Tukey, 1977)

- Statistical analysis provided metaknowledge used in the ontologies

Task-based evaluation (Obrst et al., 2007) to check accuracy and explainability of ontologies. Accurate data extraction using DL queries and SPARQL

- MoJ class structure unambiguous and explainable
- Transparency of input and output improved

Limitations

- The MoJ ontology was populated with dummy data because real were unavailable.
- Correlation features are separate from object and data properties storing criminal justice data

Conclusions

Bias can be introduced through:

- Feature selection
- Performance metrics
- Data

Transparency is key to ethics and fairness

Ontology is a credible solution to mitigate bias with:

- Transparency of input
- Transparency of output

Recommendations

MoJ ontology to be industrialised by:

- Review design with subject matter expert
- Extend domain beyond recidivism to cover entirely of criminal justice in England and Wales
- Populate with real data to validate potential biases with existing tools (OGRS3)

Include dynamic features in the next iteration of OGRS

Review semantic web with W3C to store metaknowledge as knowledge

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016) Machine Bias. Available from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. [Accessed 29 April 2024].

Bārzdiņš, J., Bārzdiņš, G., Čerāns, K., Liepiņš, R. & Sproģis, A. (2010) 'UML style graphical notation and editor for OWL 2', *Perspectives in Business Informatics Research: 9th International Conference*, Rostock Germany, 29 September – 1 October 1. Berlin Heidelberg: Springer.102-114.

BCS (2022) Code of Conduct for BCS Members. Available from: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 15 July 2024].

Biddle, J.B. (2022) On predicting recidivism: epistemic risk, tradeoffs, and values in machine learning. *Canadian Journal of Philosophy*, 52(3): 321-341. https://doi.org/10.1017/can.2020.27

Borden, H.G. (1928) Factors for predicting parole success. *Journal of the American Institute of Criminal Law and Criminology*, 19(3): 328-336.

Bushway, S.D. & Piehl, A.M. (2007) The inextricable link between age and criminal history in sentencing. *Crime & Delinquency*, 53(1): 156-183. https://doi.org/10.1177/0011128706294444

Caton, S. & Haas, C. (2020) Fairness in machine learning: A survey. *ACM Computing Surveys*, 56(7): 1-38. https://doi.org/10.1145/3616865

Curtis, J. (2018) On using machine learning to predict recidivism. Ph.D. thesis, Texas Tech University. Available from: https://ttu-ir.tdl.org/server/api/core/bitstreams/8e745777-200b-45d2-94a1-84355598d2ba/content [Accessed 22 April 2024].

Equivant (2019) Practitioner's Guide to COMPAS Core. Available from: https://www.equivant.com/wp-content/uploads/Practitioners-Guide-to-COMPAS-Core-040419.pdf [Accessed 30 April 2024].

HM Prison & Probation Service (2023) Risk of Serious Harm Guidance 2020 v3. London: HM Prison & Probation Service. Available from: https://assets.publishing.service.gov.uk/media/652cf8c9697260000dccf834/Risk_of_Serious_Harm_Guidance_v3.pdf [Accessed 17 April 2024].

Kim, J.H. and Choi, I. (2021) Choosing the level of significance: A decision-theoretic approach. *Abacus*, 57(1): 27-71.

References

Kleiman, M., Ostrom, B.J. & Cheesman, F.L. (2007) Using risk assessment to inform sentencing decisions for nonviolent offenders in Virginia. *Crime & Delinquency*, 53(1): 106-132. https://doi.org/10.1177/0011128706294442

Kleinberg, J., Mullainathan, S. and Raghavan, M. (2016) Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*. https://doi.org/10.48550/arXiv.1609.05807

Kovalchuk, O., Karpinski, M., Banakh, S., Kasianchuk, M., Shevchuk, R. & Zagorodna, N. (2023) Prediction machine learning models on propensity convicts to criminal recidivism. *Information*, 14(3): 1-15. https://doi.org/10.3390/info14030161

Lin, Z.J., Jung, J., Goel, S. & Skeem, J. (2020) The limits of human predictions of recidivism. *Science advances*, 6(7): 1-8.

Ministry of Justice (2020) Ministry of Justice: Data First. Available from: https://www.gov.uk/guidance/ministry-of-justice-data-first [Accessed 18 March 2024].

Nafekh, M. & Motiuk, L.L. (2002) *The statistical information on recidivism, revised 1 (SIR-R1) scale: a psychometric examination*. Ottawa, Ontario: Correctional Service of Canada, Research Branch.

National Institute of Justice (N.D.) Recidivism Forecasting Challenge. Available from: https://nij.ojp.gov/funding/recidivism-forecasting-challenge [Accessed 18 March 2024].

Obrst, L., Ceusters, W., Mani, I., Ray, S. & Smith, B. (2007) The evaluation of ontologies: Toward improved semantic interoperability. *Semantic web: Revolutionizing knowledge discovery in the life sciences*: 139-158.

Rudin, C. (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5): 206-215.

Skeem, J., Monahan, J. & Lowenkamp, C. (2016) Gender, risk assessment, and sanctioning: The cost of treating women like men. *Law and human behavior*, 40(5): 580-593. https://doi.org/10.1037/lhb0000206

Squadrone, L., Croce, D. & Basili, R. (2022) ‘Ethics by design for intelligent and sustainable adaptive systems’, *International Conference of the Italian Association for Artificial Intelligence*. Udine, Italy, 28 November – 2 December. Cham: Springer International Publishing. 154-167.

References

Stevenson, M.T. & Slobogin, C. (2018) Algorithmic risk assessments and the double-edged sword of youth. *Behavioral sciences & the law*, 36(5): 638-656. https://doi.org/10.1002/bsl.2384

Tollenaar, N. & van der Heijden, P.G. (2013) Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 176(2): 565-584. https://doi.org/10.1111/j.1467-985X.2012.01056.x

Tukey, J.W. (1977) *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Walmsley, J. (2021) Artificial intelligence and the value of transparency. *AI & Society*, 36(2): 585-595.

Wang, P., Mathieu, R., Ke, J. & Cai, H.J. (2010) 'Predicting criminal recidivism with support vector machine'. International Conference on Management and Service Science. Wuhan, China, 24-26 August. IEEE. 1-9. https://doi.org/10.1109/ICMSS.2010.5575352

Xiao, C., Ye, J., Esteves, R.M. & Rong, C. (2016) Using Spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency and Computation: Practice and Experience*, 28(14): 3866-3878. https://doi.org/10.1002/cpe.3745

Zeng, J., Ustun, B. & Rudin, C. (2017) Interpretable classification models for recidivism prediction. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180(3): 689-722. https://doi.org/10.1111/rssa.12227

Thank You

Artefacts available at:

https://github.com/feaviolp/msc-project/