

Dossier de conformité — données d'entraînement (AI Act, article 10)

Squelette généré automatiquement à partir de l'analyse du jeu de données. Les blocs « à compléter par l'expert » nécessitent une revue humaine (provenance, jugement de conformité, mesures d'atténuation).

1. Identité & finalité

- **Système / modèle** : _____
- **Version du dossier** : _____
- **Date d'analyse** : _____
- **Jeu de données** : src
- **Finalité prévue** : _____

À compléter par l'expert — finalité prévue et ce que les données sont censées représenter

2. Provenance & base légale

Origine de chaque source, licences/contrats, et — pour les données personnelles — finalité initiale et base légale RGPD.

(!) **Données personnelles potentielles détectées** : telephone — une base légale RGPD est requise pour ces colonnes.

À compléter par l'expert — provenance par source (flux séparé §16), licences, base légale RGPD

3. Composition (rempli automatiquement)

- **Volume** : 1000 lignes, 21 colonnes.
- **Variables** :

Colonne	Type
status	VARCHAR
duration	BIGINT
credit_history	VARCHAR
purpose	VARCHAR
amount	BIGINT
savings	VARCHAR
employment_duration	VARCHAR
installment_rate	BIGINT
personal_status_sex	VARCHAR
other_debtors	VARCHAR
present_residence	BIGINT
property	VARCHAR
age	BIGINT

other_installment_plans	VARCHAR
housing	VARCHAR
number_credits	BIGINT
job	VARCHAR
people_liable	BIGINT
telephone	BOOLEAN
foreign_worker	BOOLEAN
credit_risk	BIGINT

• **Attributs sensibles déclarés** : status, credit_history, purpose, savings, employment_duration, personal_status_sex, other_debtors, property, age, other_installment_plans, housing, job, foreign_worker.

• **Colonnes PII détectées** : telephone.

À compléter par l'expert — périmètre géographique / contextuel / comportemental

4. Préparation (observations automatiques)

À compléter par l'expert — journal des transformations : collecte, nettoyage, étiquetage, enrichissement, agrégation

5. Qualité (rempli automatiquement)

• **Exactitude / aberrations** : duration (1).

À compléter par l'expert — représentativité et pertinence vis-à-vis de la finalité

6. Biais

Analyse automatique

Caractéristique	Mesure	Cible	Verdict
Complétude	100.0 %	"e 95.0 %	conforme
Unicité	100.0 %	"e 99.0 %	conforme
Validité	99.7 %	"e 98.0 %	conforme

status — « credit_risk » : **disparate impact** (ratio) 1.74 · **parité statistique** (écart) 37.6 % (!)

Groupe	Effectif	Taux de « credit_risk »
no checking account	394	88.3 %
... >= 200 DM / salary for at least 1 year	63	77.8 %
0 <= ... < 200 DM	269	61.0 %
... < 100 DM	274	50.7 %

credit_history — « credit_risk » : **disparate impact** (ratio) 2.21 · **parité statistique** (écart) 45.4 % (!)

Groupe	Effectif	Taux de « credit_risk »
critical account/other credits existing	293	82.9 %
delay in paying off in the past	88	68.2 %
existing credits paid back duly till now	530	68.1 %
all credits at this bank paid back duly	49	42.9 %
no credits taken/all credits paid back duly	40	37.5 %

purpose — « credit_risk » : **disparate impact** (ratio) 1.59 · **parité statistique** (écart) 32.9 % (!)

(!) Disparité **non robuste** : plus petit groupe n=9 — écart non significatif ou porté par un petit effectif.
 À ne pas interpréter comme un biais avéré._

Groupe	Effectif	Taux de « credit_risk »
business	9	88.9 %
car (used)	103	83.5 %
domestic appliances	280	77.9 %
radio/television	181	68.0 %
repairs	12	66.7 %
others	97	64.9 %
education	22	63.6 %
car (new)	234	62.0 %
furniture/equipment	12	58.3 %
retraining	50	56.0 %

savings — « credit_risk » : **disparate impact** (ratio) 1.37 · **parité statistique** (écart) 23.5 % (!)

Groupe	Effectif	Taux de « credit_risk »
... >= 1000 DM	48	87.5 %
500 <= ... < 1000 DM	63	82.5 %
unknown/no savings account	183	82.5 %
100 <= ... < 500 DM	103	67.0 %
... < 100 DM	603	64.0 %

employment_duration — « credit_risk » : **disparate impact** (ratio) 1.31 · **parité statistique** (écart) 18.3 % (!)

Groupe	Effectif	Taux de « credit_risk »
4 <= ... < 7 years	174	77.6 %

... >= 7 years	253	74.7 %
1 <= ... < 4 years	339	69.3 %
unemployed	62	62.9 %
... < 1 year	172	59.3 %

personal_status_sex — « credit_risk » : **disparate impact** (ratio) 1.22 · **parité statistique** (écart) 13.4 % (!)

Groupe	Effectif	Taux de « credit_risk »
male : single	548	73.4 %
male : married/widowed	92	72.8 %
female : divorced/separated/married	310	64.8 %
male : divorced/separated	50	60.0 %

other_debtors — « credit_risk » : **disparate impact** (ratio) 1.44 · **parité statistique** (écart) 24.7 % (!)

Groupe	Effectif	Taux de « credit_risk »
guarantor	52	80.8 %
none	907	70.0 %
co-applicant	41	56.1 %

property — « credit_risk » : **disparate impact** (ratio) 1.39 · **parité statistique** (écart) 22.2 % (!)

Groupe	Effectif	Taux de « credit_risk »
real estate	282	78.7 %
building society savings agreement/ life insurance	232	69.4 %
car or other	332	69.3 %
unknown/no property	154	56.5 %

other_installment_plans — « credit_risk » : **disparate impact** (ratio) 1.23 · **parité statistique** (écart) 13.5 % (!)

Groupe	Effectif	Taux de « credit_risk »
none	814	72.5 %
stores	47	59.6 %
bank	139	59.0 %

housing — « credit_risk » : **disparate impact** (ratio) 1.25 · **parité statistique** (écart) 14.6 % (!)

Groupe	Effectif	Taux de « credit_risk »
own	713	73.9 %

rent	179	60.9 %
for free	108	59.3 %

job — « credit_risk » : **disparate impact** (ratio) 1.10 · **parité statistique** (écart) 6.5 % (!)

(!) Disparité **non robuste** : plus petit groupe n=148 — écart non significatif ou porté par un petit effectif. À ne pas interpréter comme un biais avéré.

Groupe	Effectif	Taux de « credit_risk »
unskilled - resident	200	72.0 %
skilled employee/official	630	70.5 %
unemployed/unskilled - non-resident	22	68.2 %
management/self-employed/highly qualified employee/officer	148	65.5 %

foreign_worker — « credit_risk » : **disparate impact** (ratio) 1.29 · **parité statistique** (écart) 19.9 % (!)

Groupe	Effectif	Taux de « credit_risk »
false	37	89.2 %
true	963	69.3 %

Jugement & mesures

À compléter par l'expert — impact sur santé/sécurité/droits fondamentaux et mesures de détection/prévention/atténuation

7. Lacunes & limites

Points détectés à examiner comme lacunes potentielles :

- [low] outliers — duration
- [low] outliers — amount
- [high] pii — telephone
- [medium] bias — status
- [high] bias — credit_history
- [low] bias — purpose

À compléter par l'expert — manques identifiés et traités ; usages hors périmètre

8. Gouvernance & traçabilité

- **Version de l'outil d'analyse** : 0.4.1
- **Empreinte du jeu (schéma + volume)** : ca33a8c44d3690a0

À compléter par l'expert — responsabilités, journal d'audit, versioning, maintenance/mise à jour