

Breast Cancer

HEDI Z
TANJONA R
XAVIER T

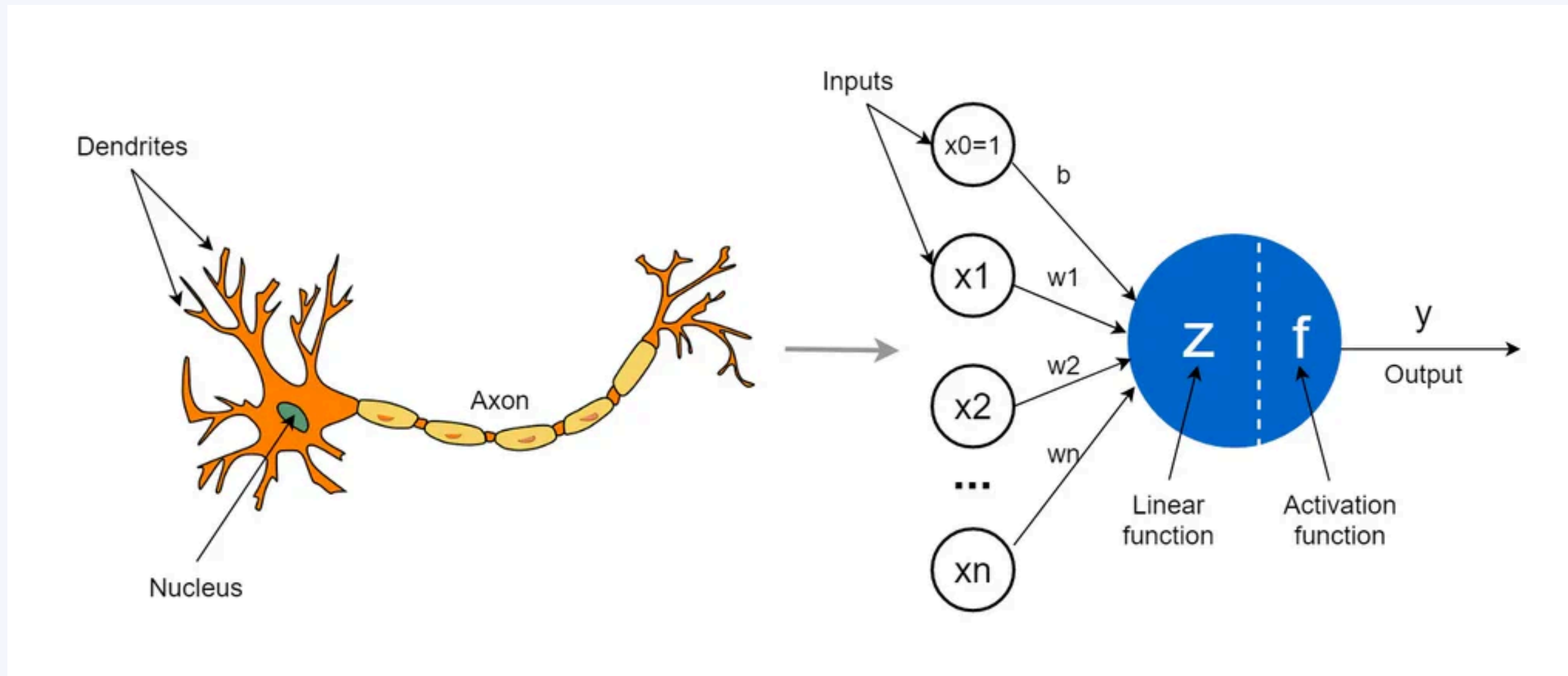


Sommaire

1. Présentation du perceptron
2. Présentation des données
3. Sélection des variables
4. Modèle de perceptron et prédiction



Modèle de perceptron



En 2023, environ 61 214 nouveaux cas de cancer du sein ont été diagnostiqués chez les femmes, marquant une augmentation annuelle de 0,3 %.

En 2021, le cancer du sein a causé environ 12 600 décès féminins, avec une diminution annuelle de 1,3 %.



Présentation des données

Forme et Taille

radius_mean : Rayon moyen des cellules.

perimeter_mean : Périmètre moyen.

area_mean : Aire moyenne.

Texture et Symétrie

texture_mean : Variabilité des teintes dans l'image des cellules.

smoothness_mean : Régularité des contours.

symmetry_mean : Degré de symétrie des cellules.

fractal_dimension_mean : Mesure de la complexité des contour

Compactité et Concavité

compactness_mean : Compacité de la cellule

concavity_mean : Gravité des concavités dans le contour.

concave_points_mean : Nombre de points concaves dans le contour.

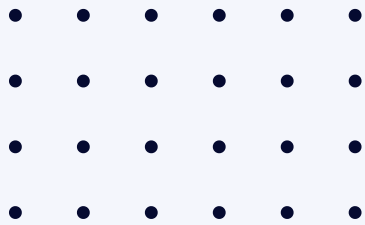
Les erreurs standard: radius_se, smoothness_se

Régions malades: radius_worst, texture_worst

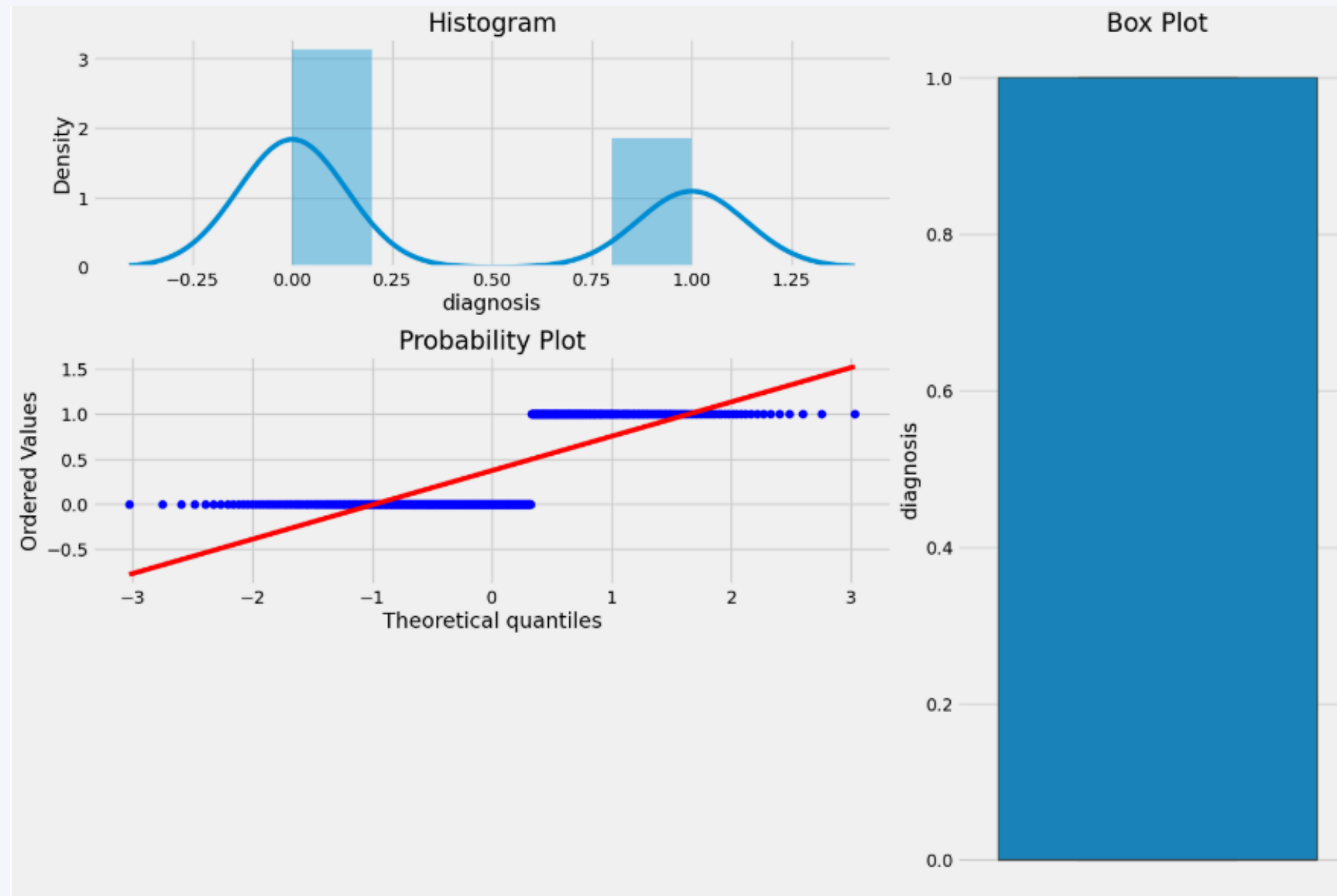


Présentation des données

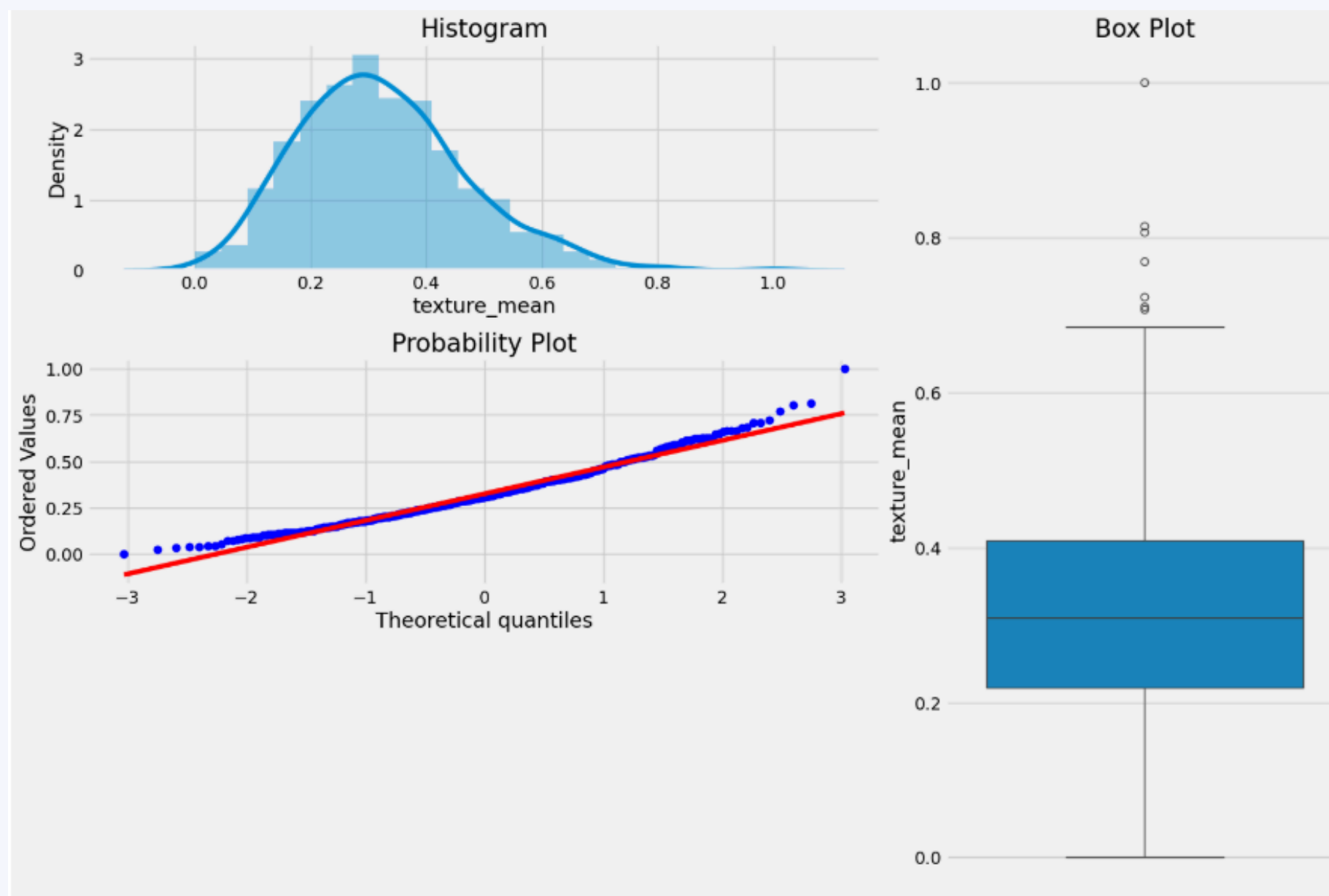
	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	symmetry_mean	...	radius_worst	texture_worst	perimeter_worst	area_worst
0	M	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710	0.2419	...	25.380	17.33	184.60	2019.0
1	M	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017	0.1812	...	24.990	23.41	158.80	1956.0
2	M	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790	0.2069	...	23.570	25.53	152.50	1709.0
3	M	11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520	0.2597	...	14.910	26.50	98.87	567.7
4	M	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430	0.1809	...	22.540	16.67	152.20	1575.0
...
564	M	21.56	22.39	142.00	1479.0	0.11100	0.11590	0.24390	0.13890	0.1726	...	25.450	26.40	166.10	2027.0
565	M	20.13	28.25	131.20	1261.0	0.09780	0.10340	0.14400	0.09791	0.1752	...	23.690	38.25	155.00	1731.0
566	M	16.60	28.08	108.30	858.1	0.08455	0.10230	0.09251	0.05302	0.1590	...	18.980	34.12	126.70	1124.0
567	M	20.60	29.33	140.10	1265.0	0.11780	0.27700	0.35140	0.15200	0.2397	...	25.740	39.42	184.60	1821.0
568	B	7.76	24.54	47.92	181.0	0.05263	0.04362	0.00000	0.00000	0.1587	...	9.456	30.37	59.16	268.6
569 rows × 31 columns															



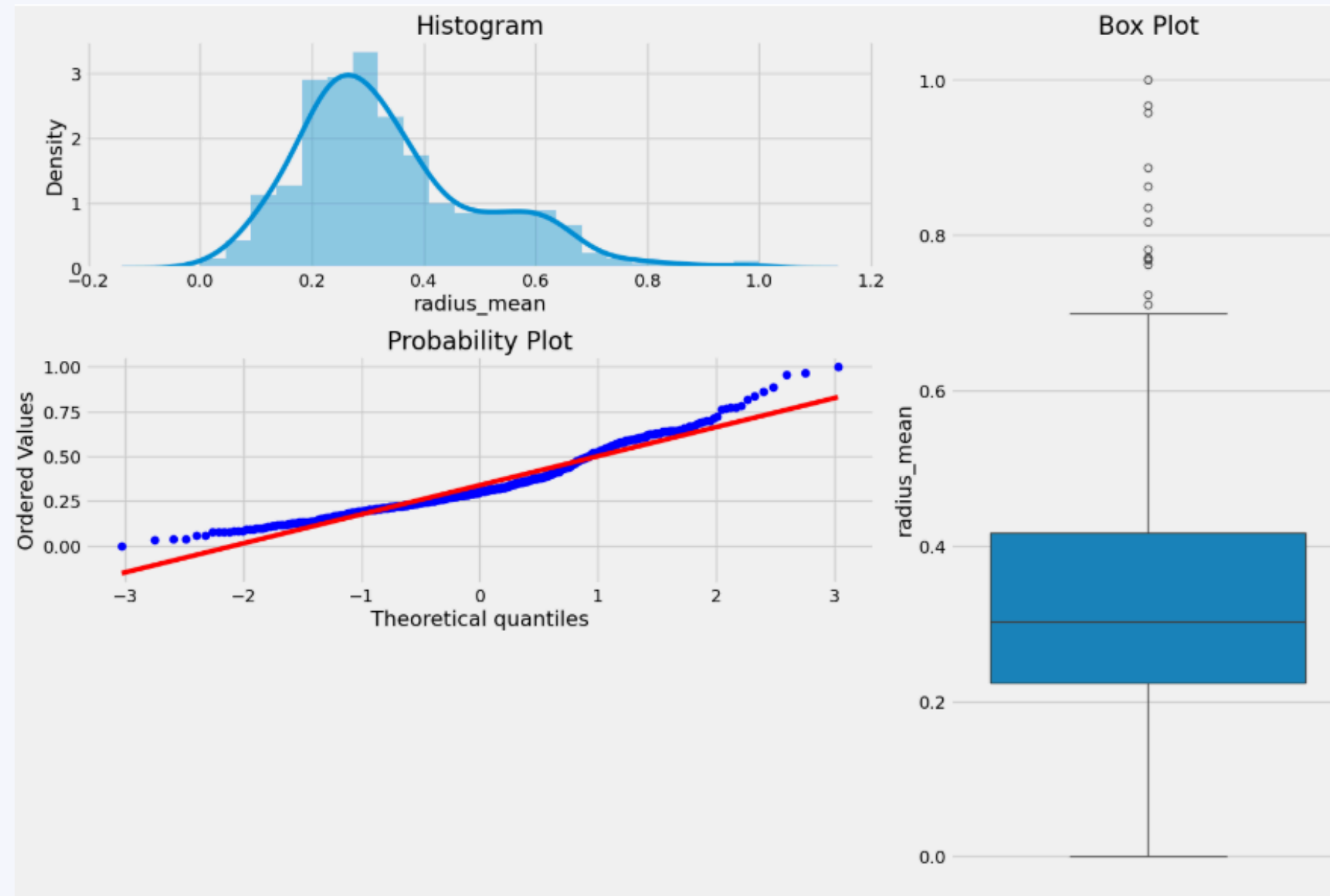
Exploration des données



Exploration des données



Exploration des données

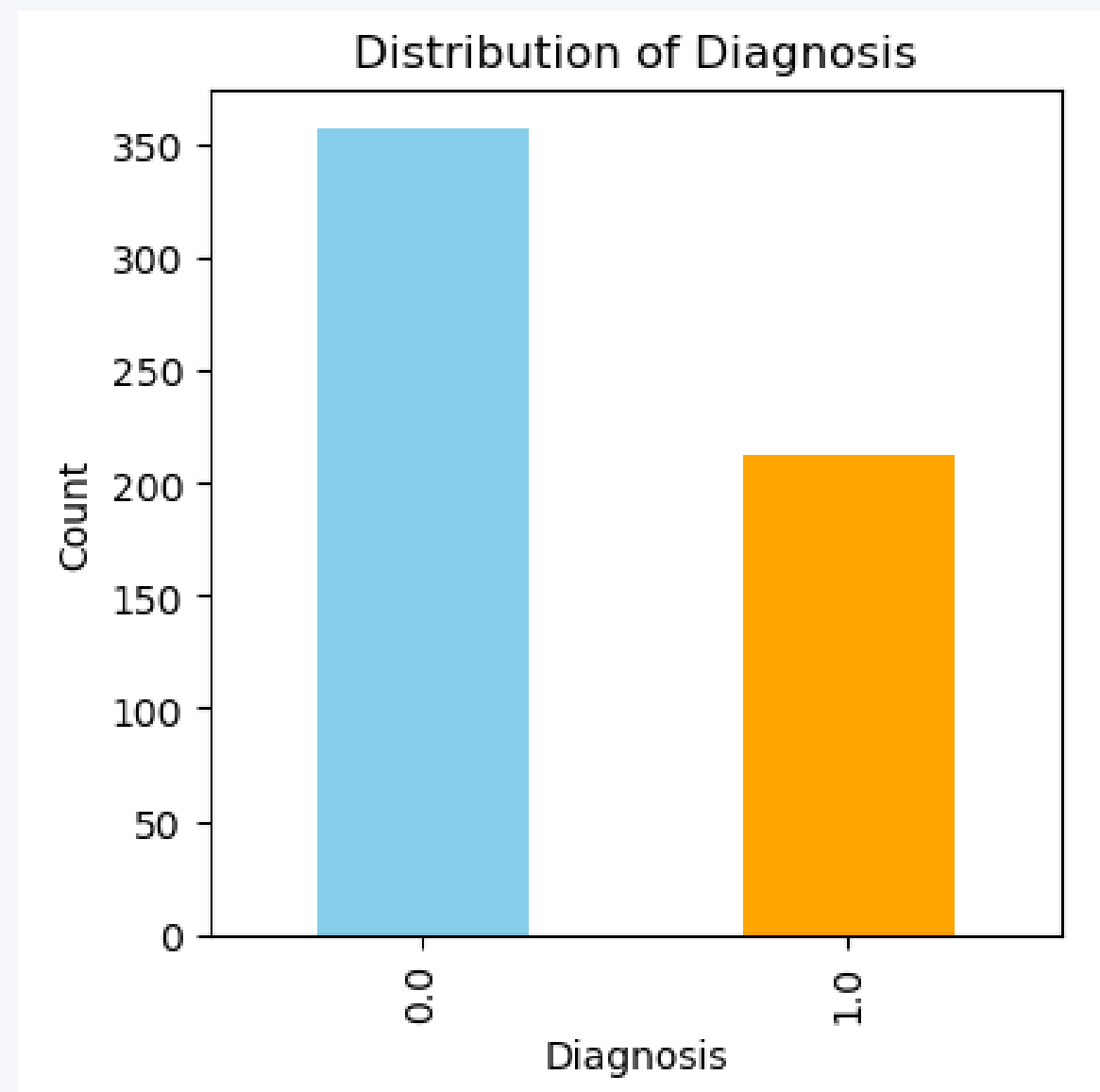


Présentation des données

BREAST CANCER WISCONSIN.

569 individus

diagnosis	0.0	1.0
count	357	212

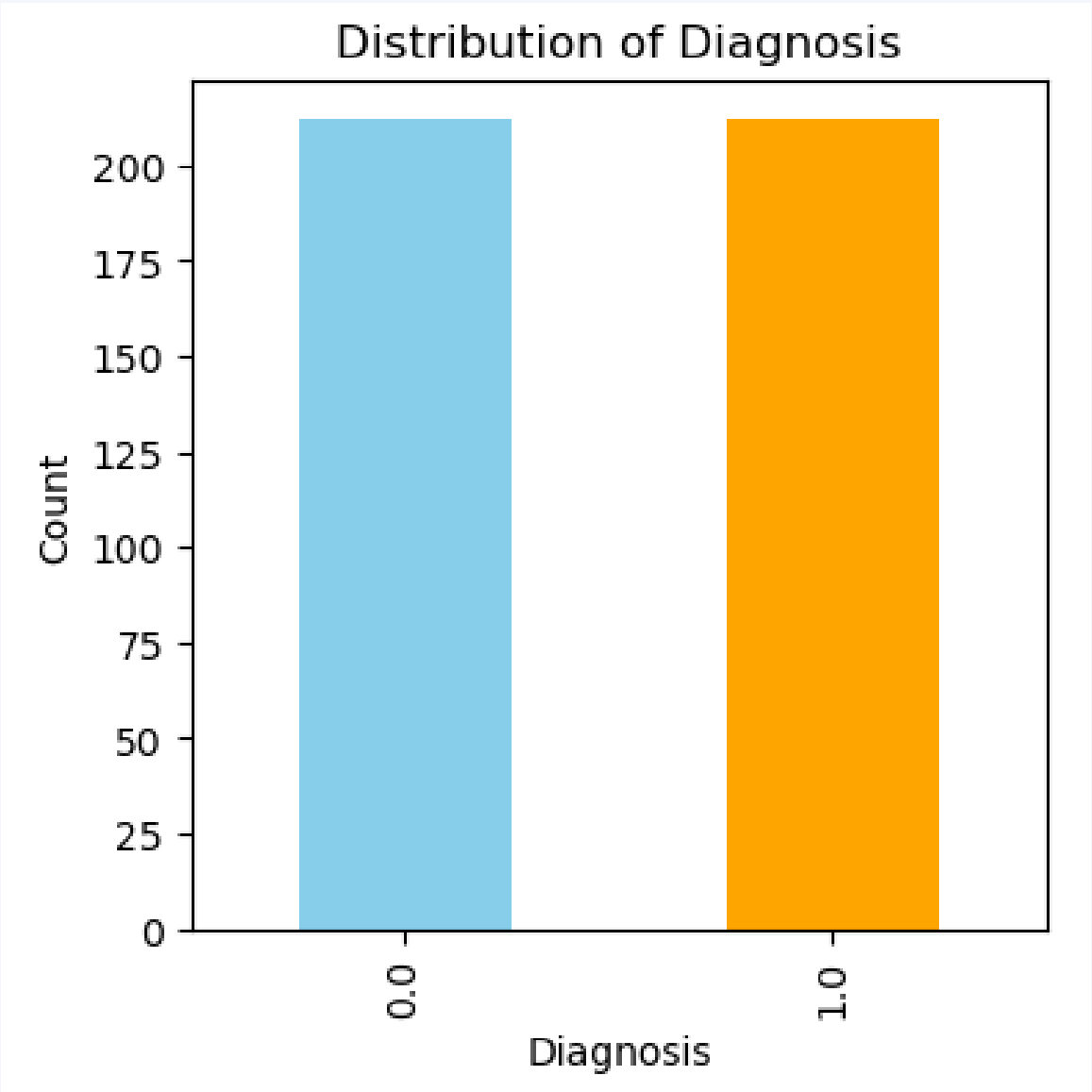


Présentation des données

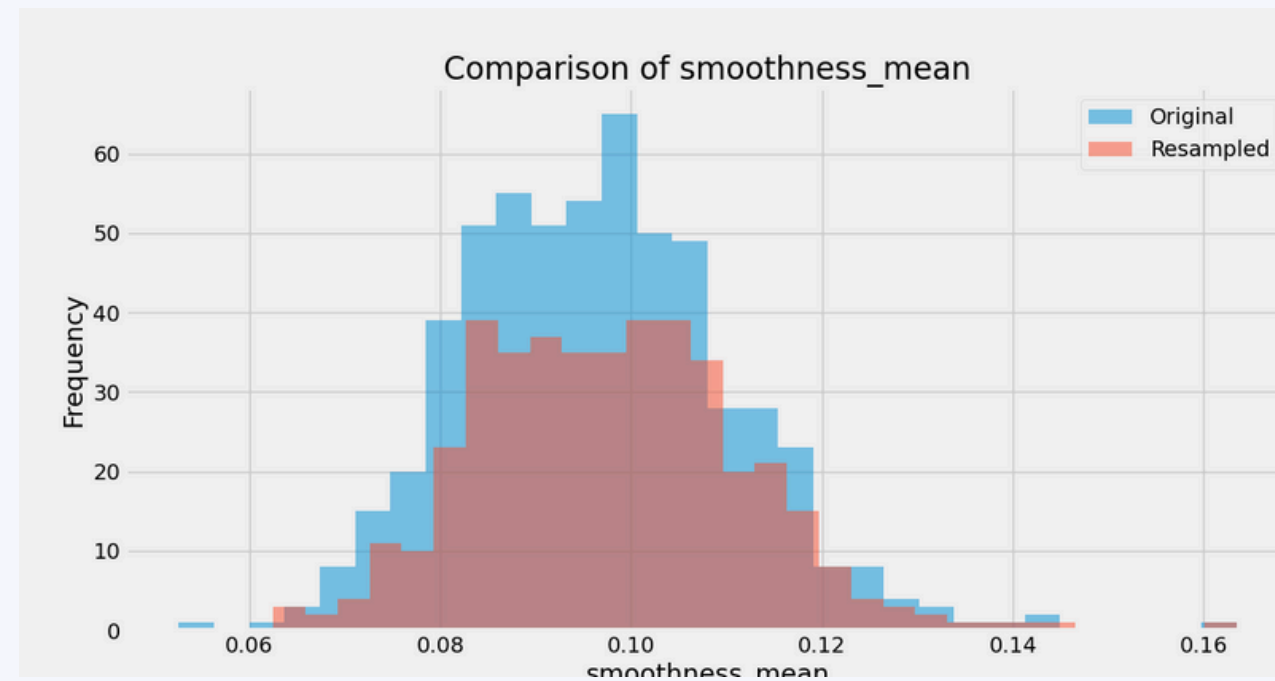
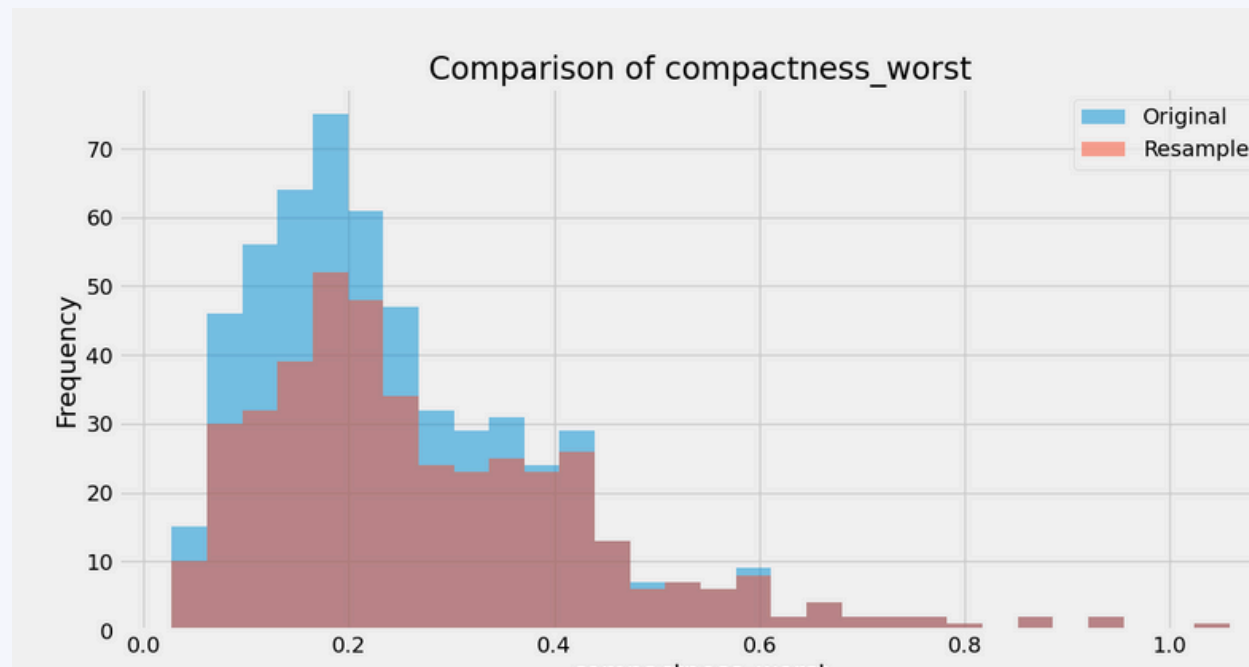
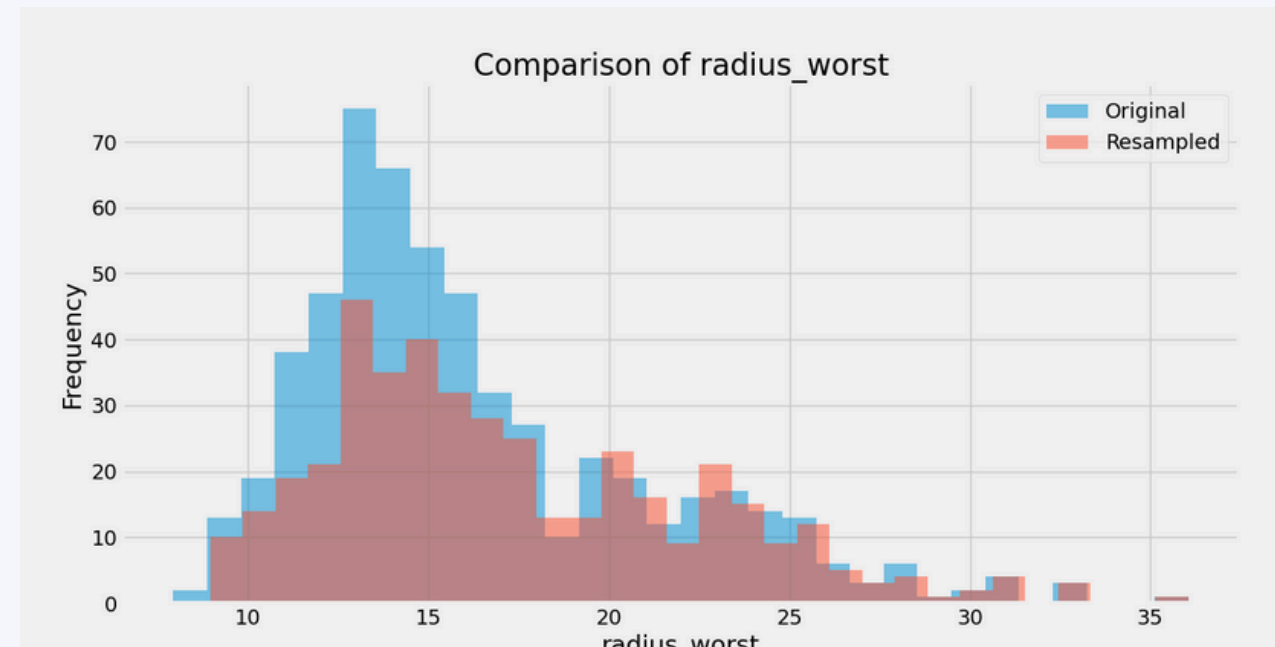
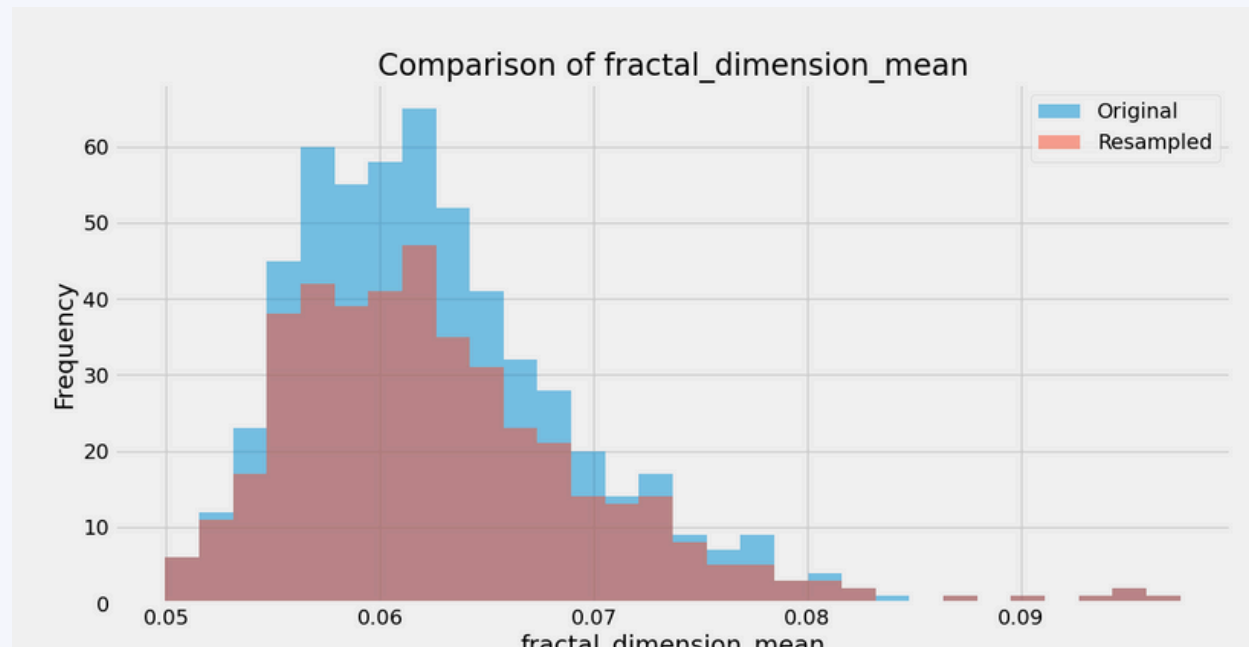
BREAST CANCER WISCONSIN.

424 individus

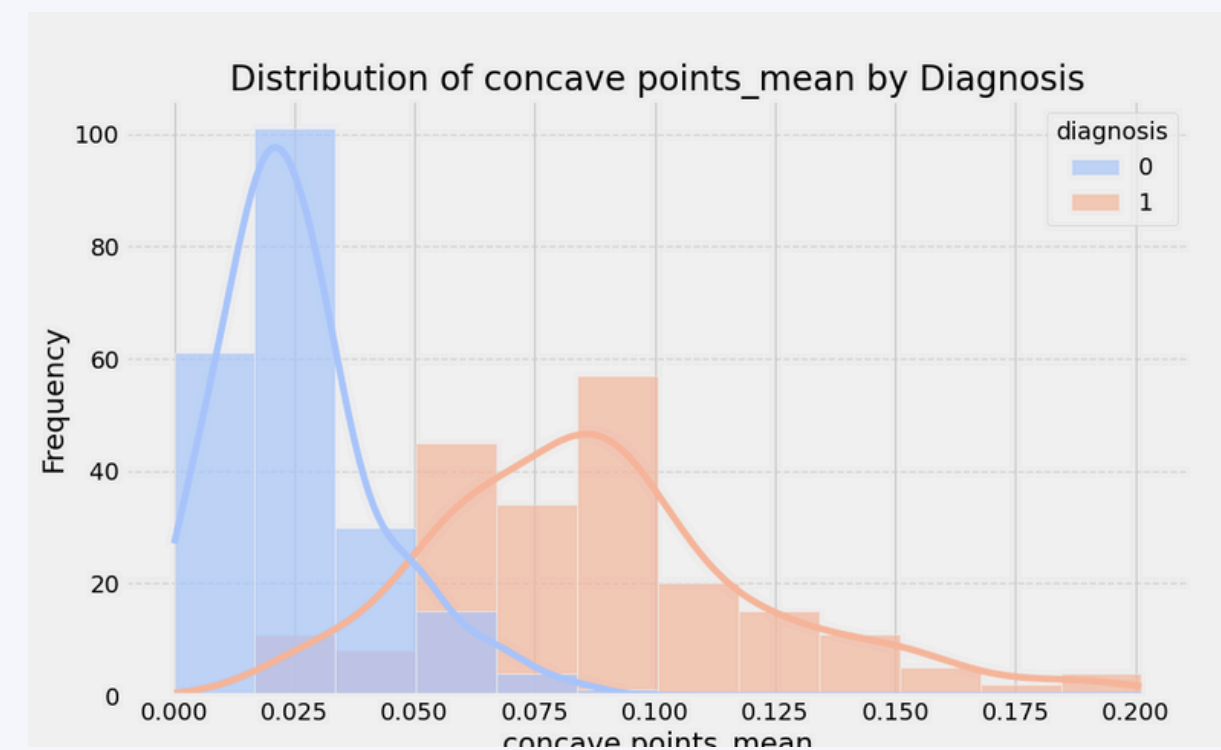
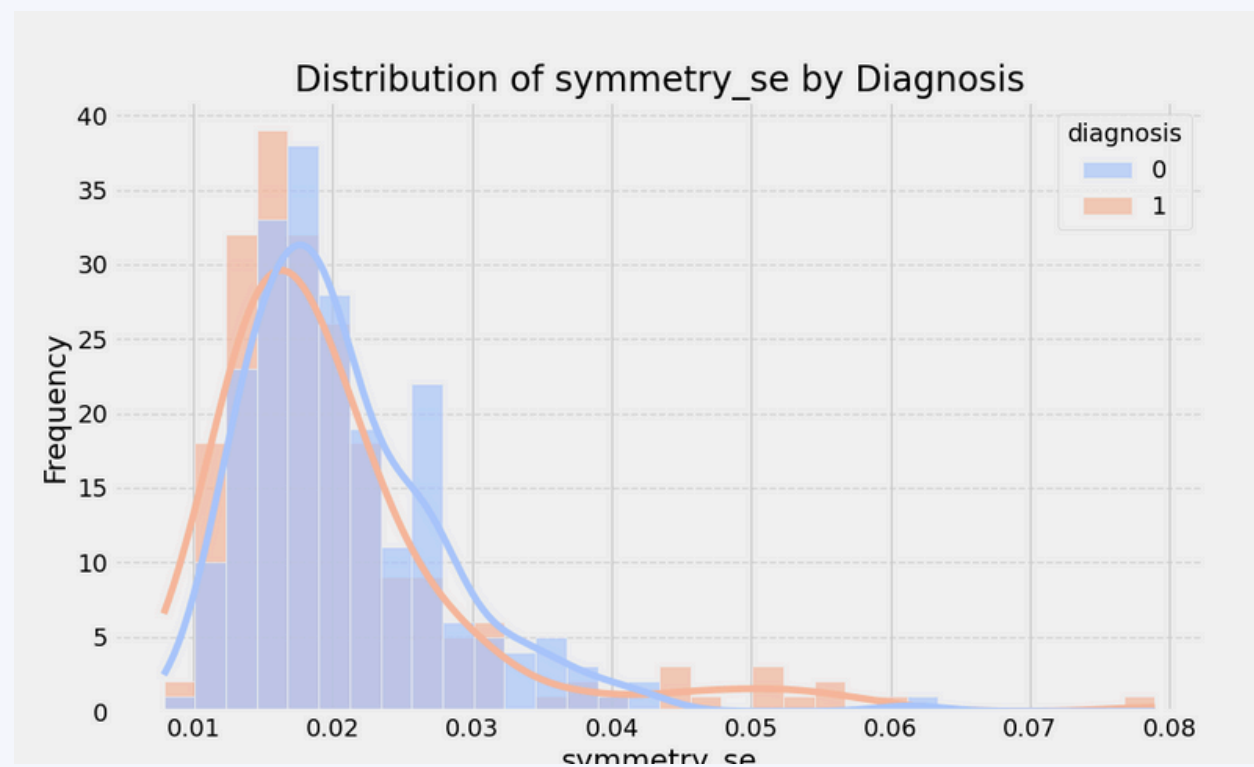
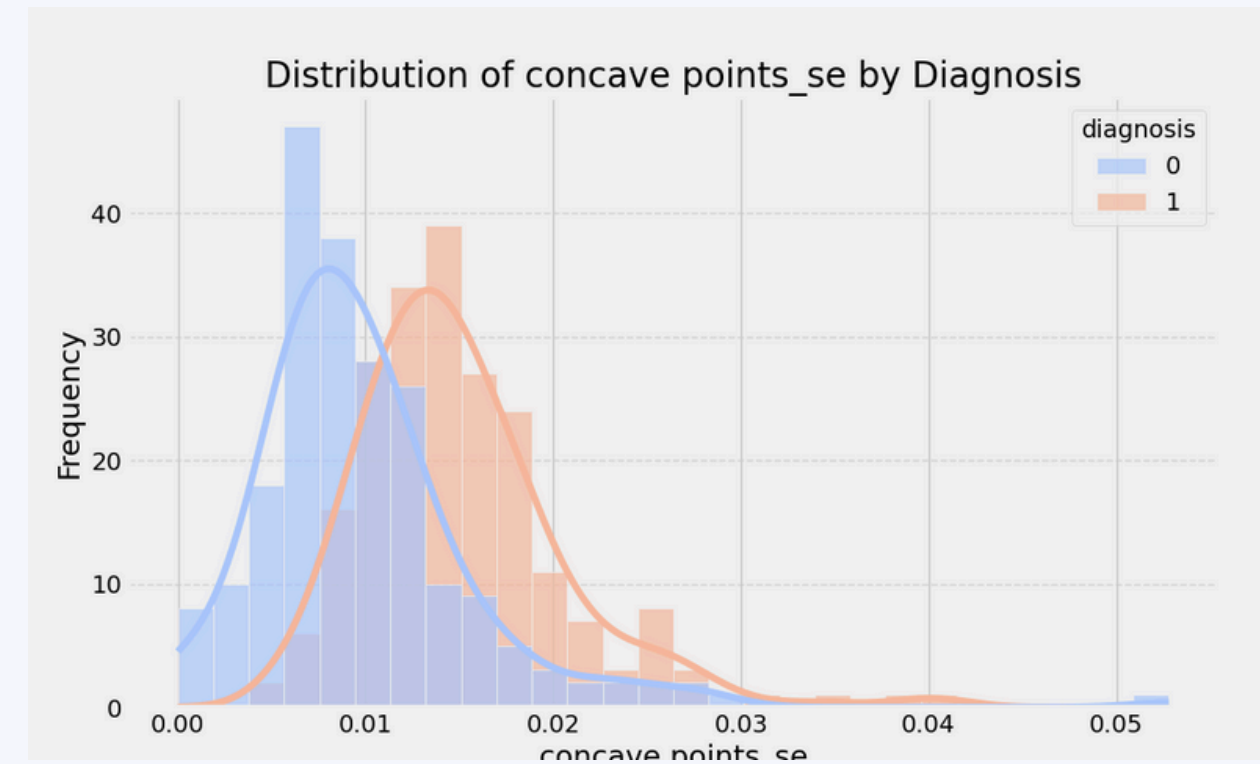
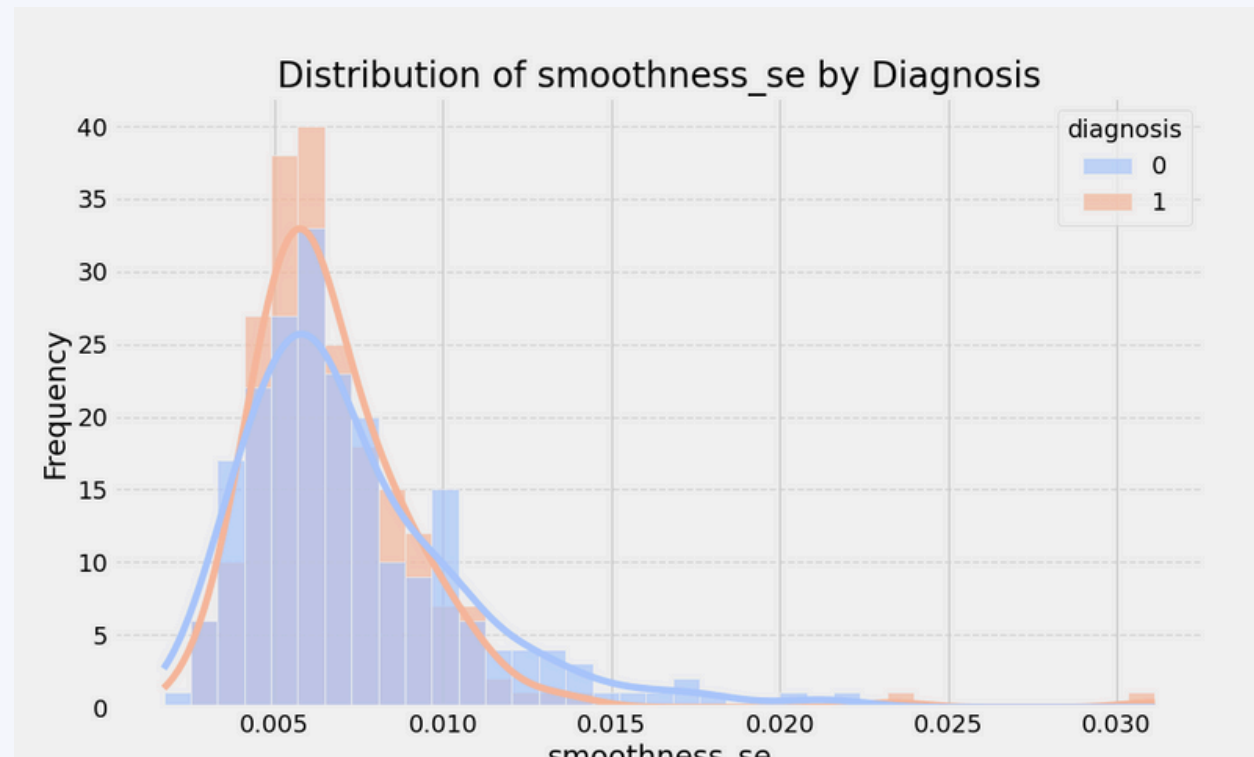
diagnosis	0.0	1.0
count	212	212



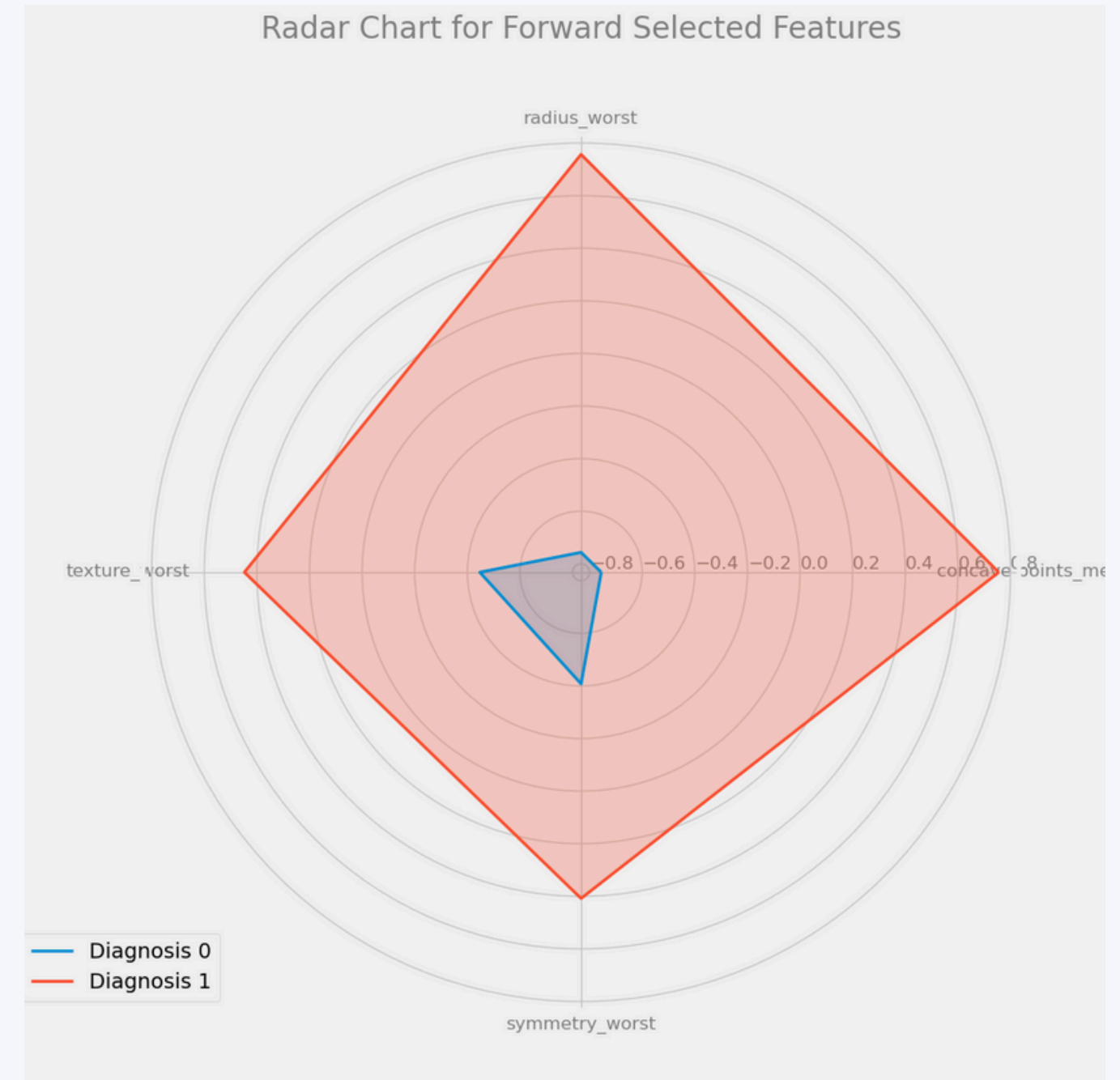
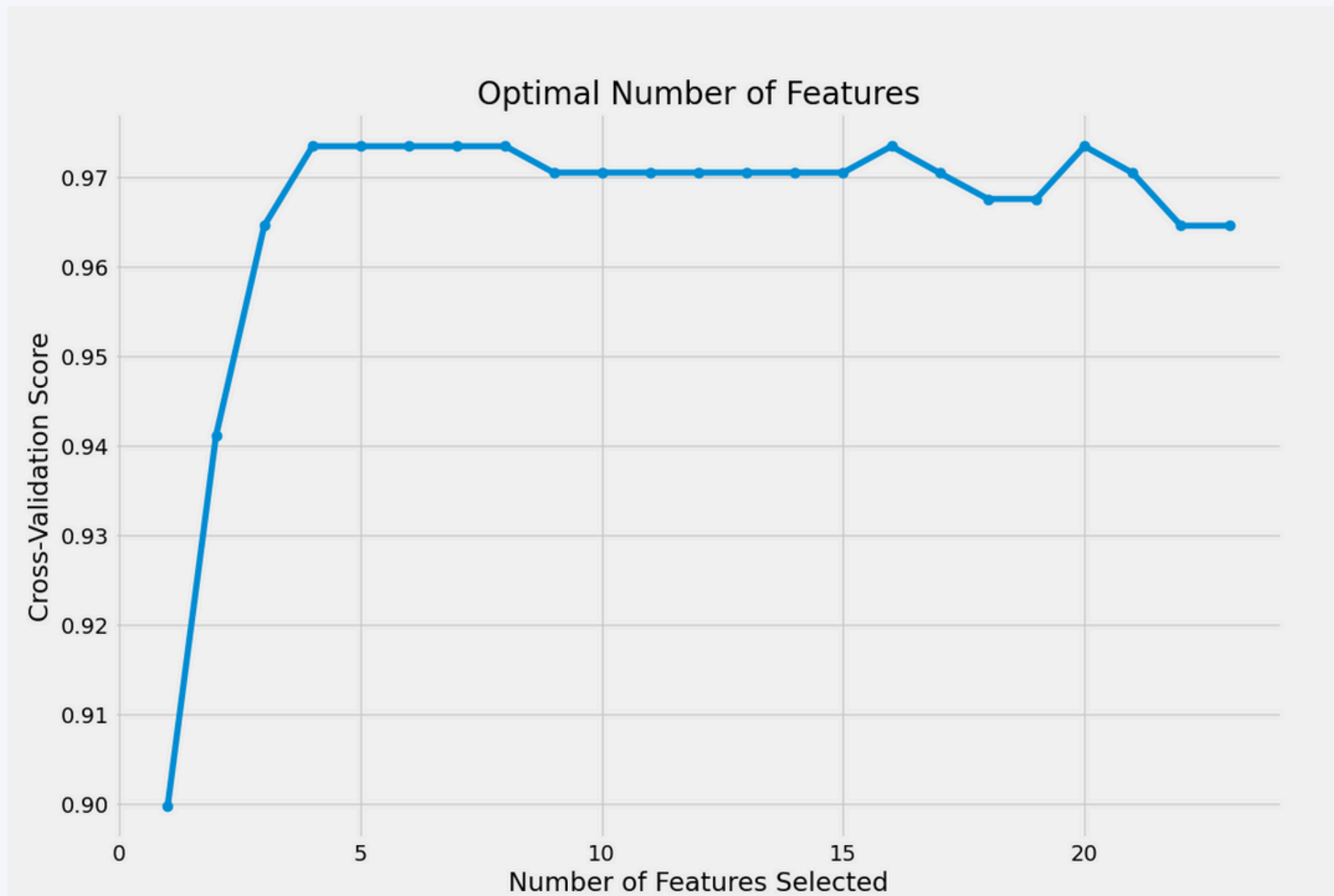
Préparation des données



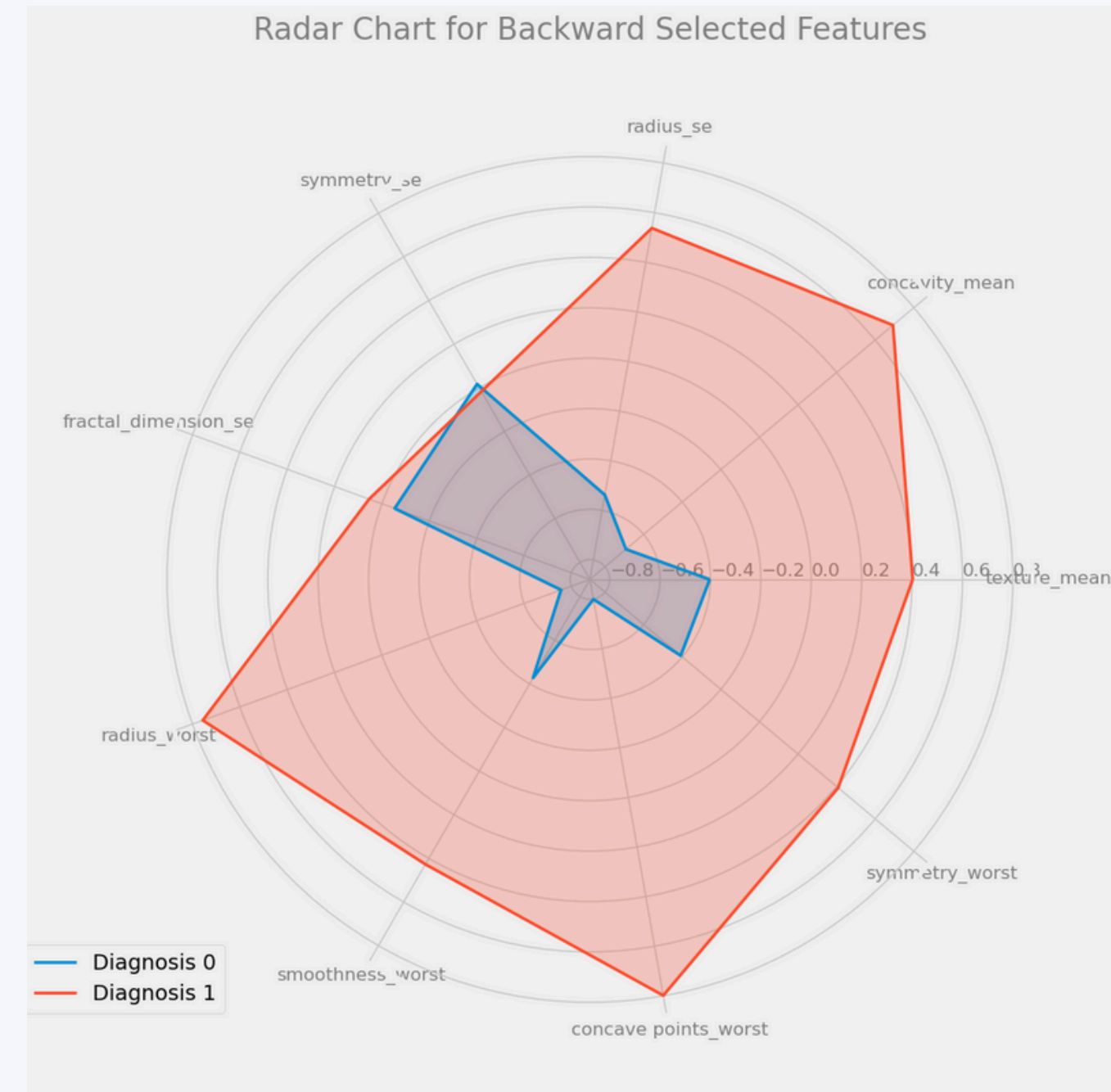
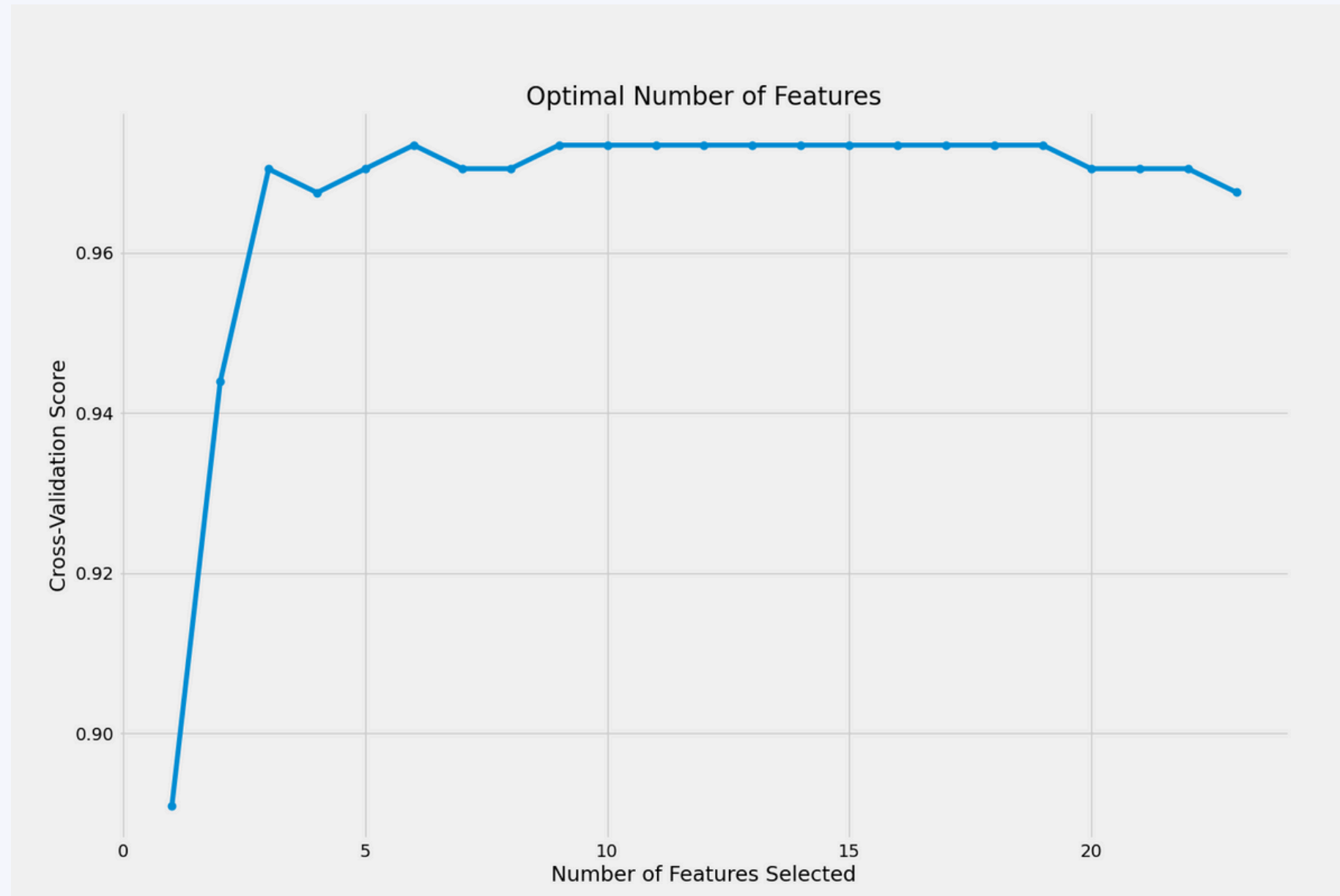
Préparation des données



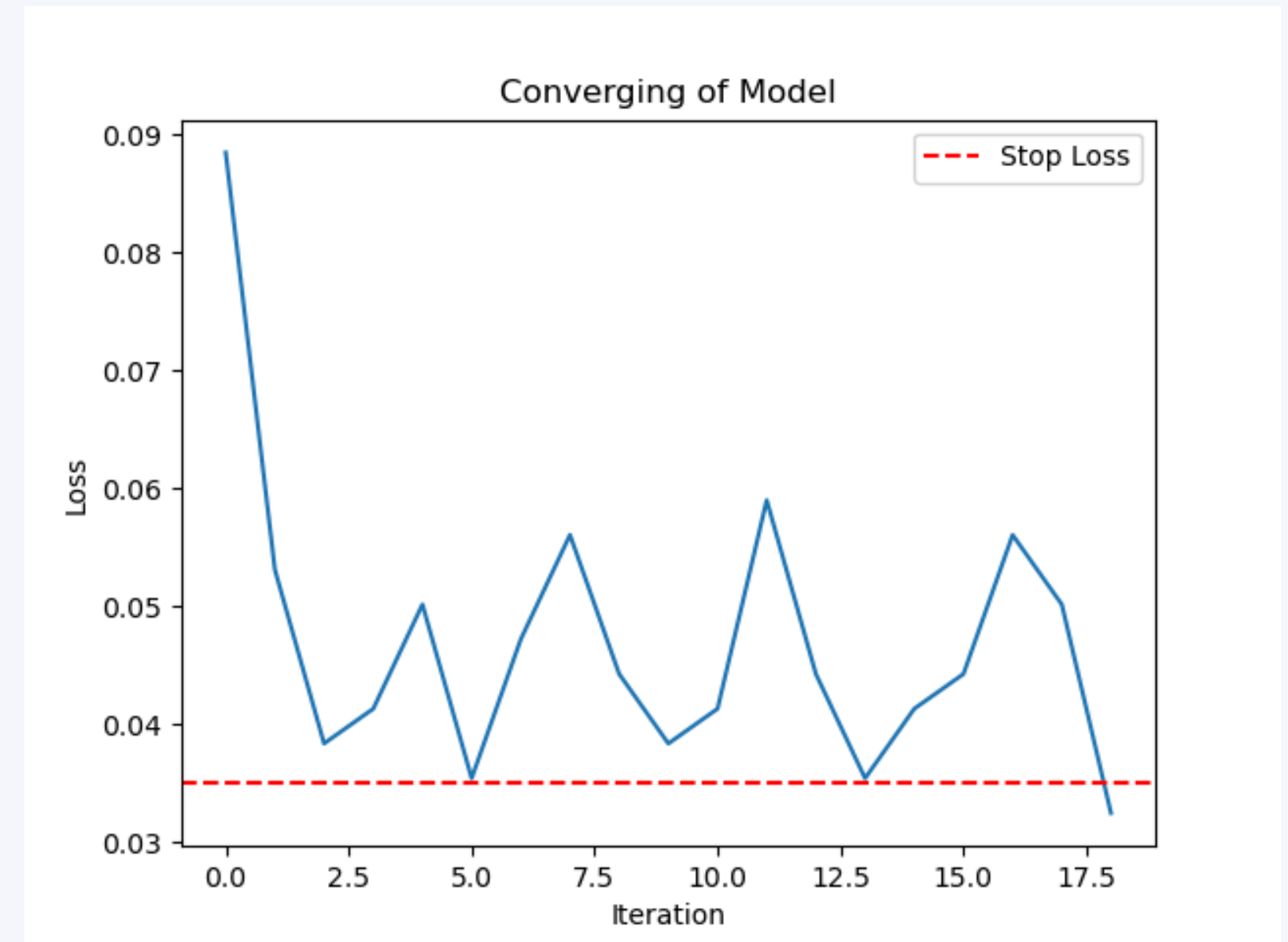
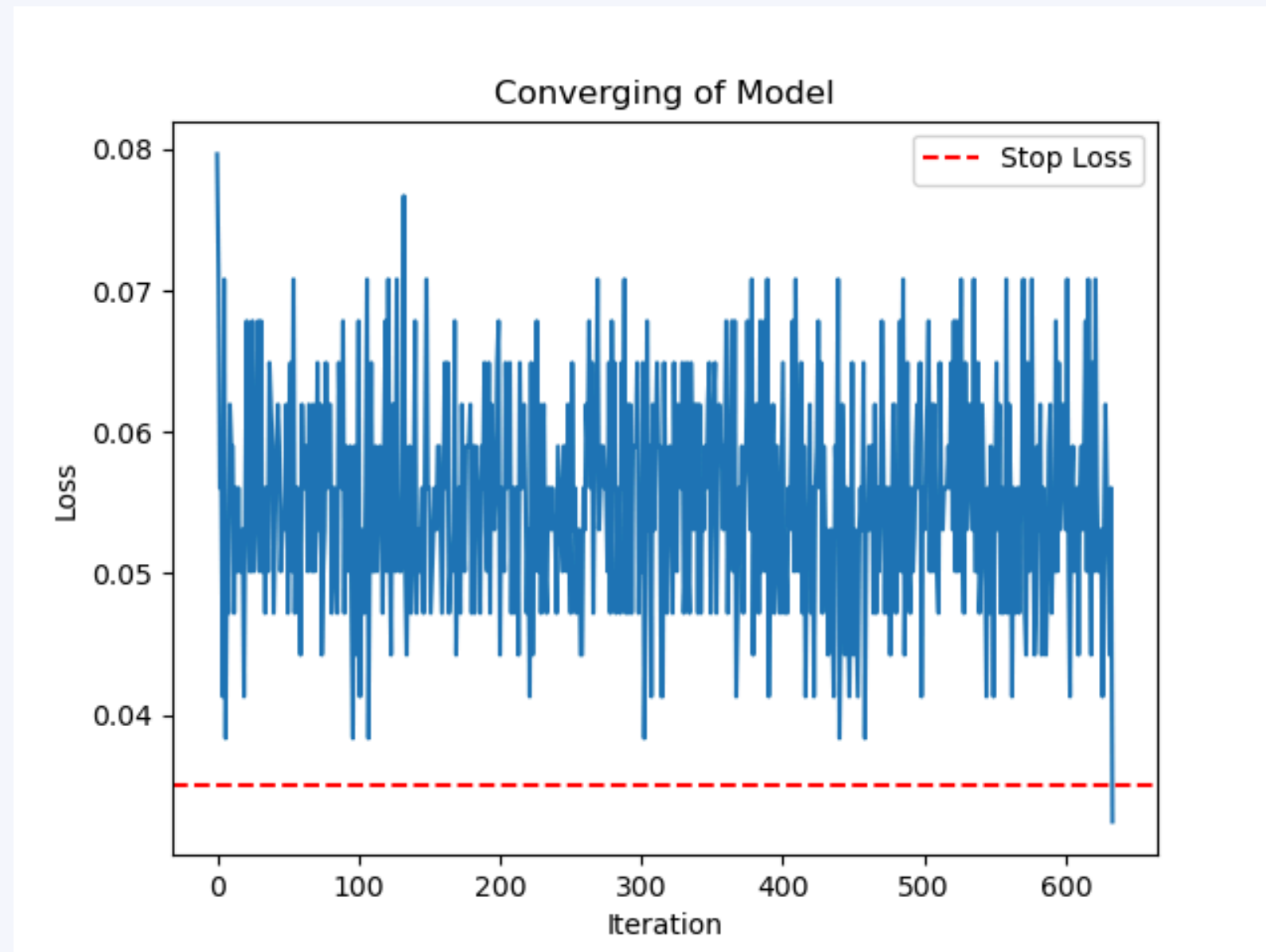
Forward selection



Backward selection

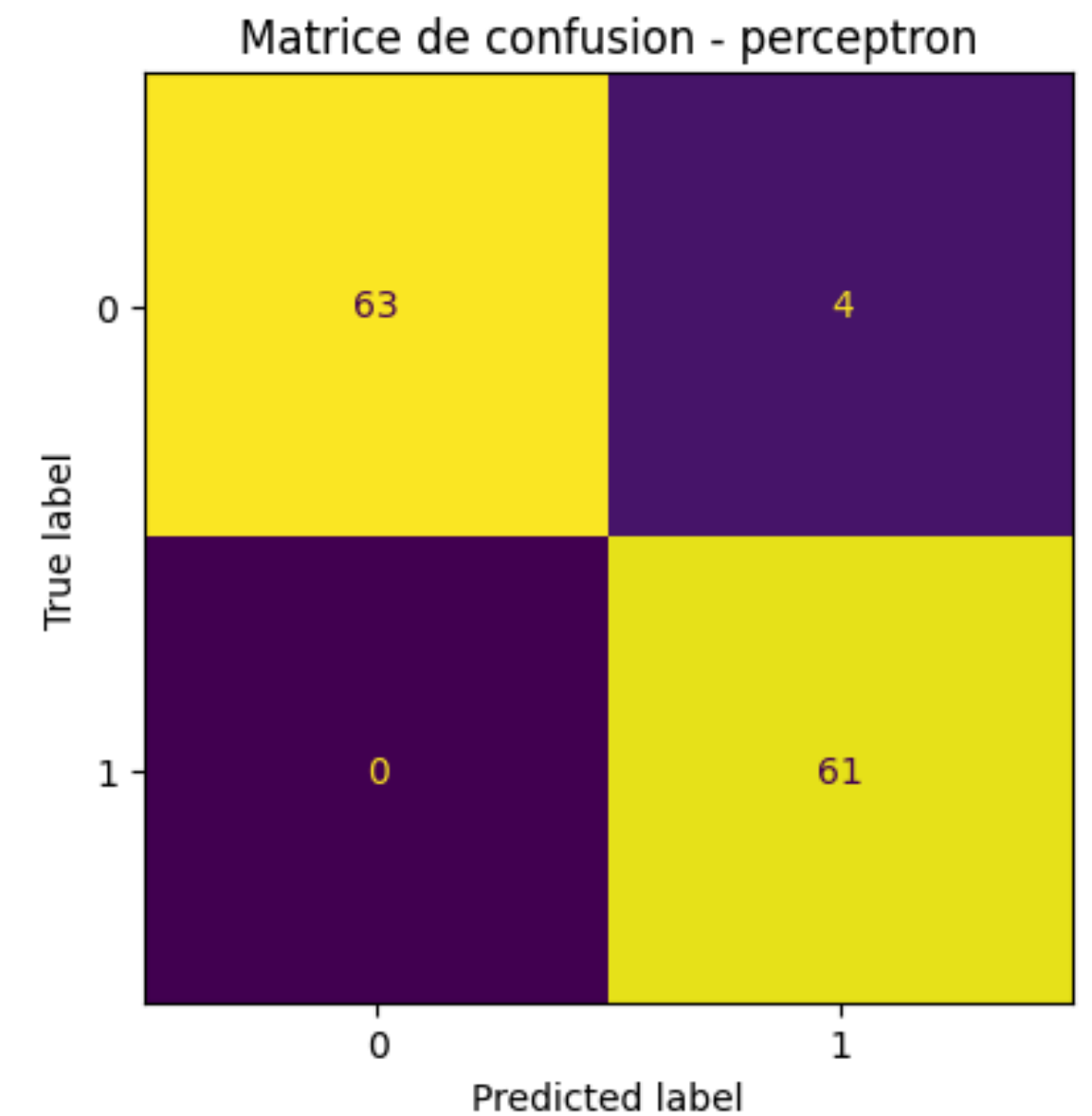
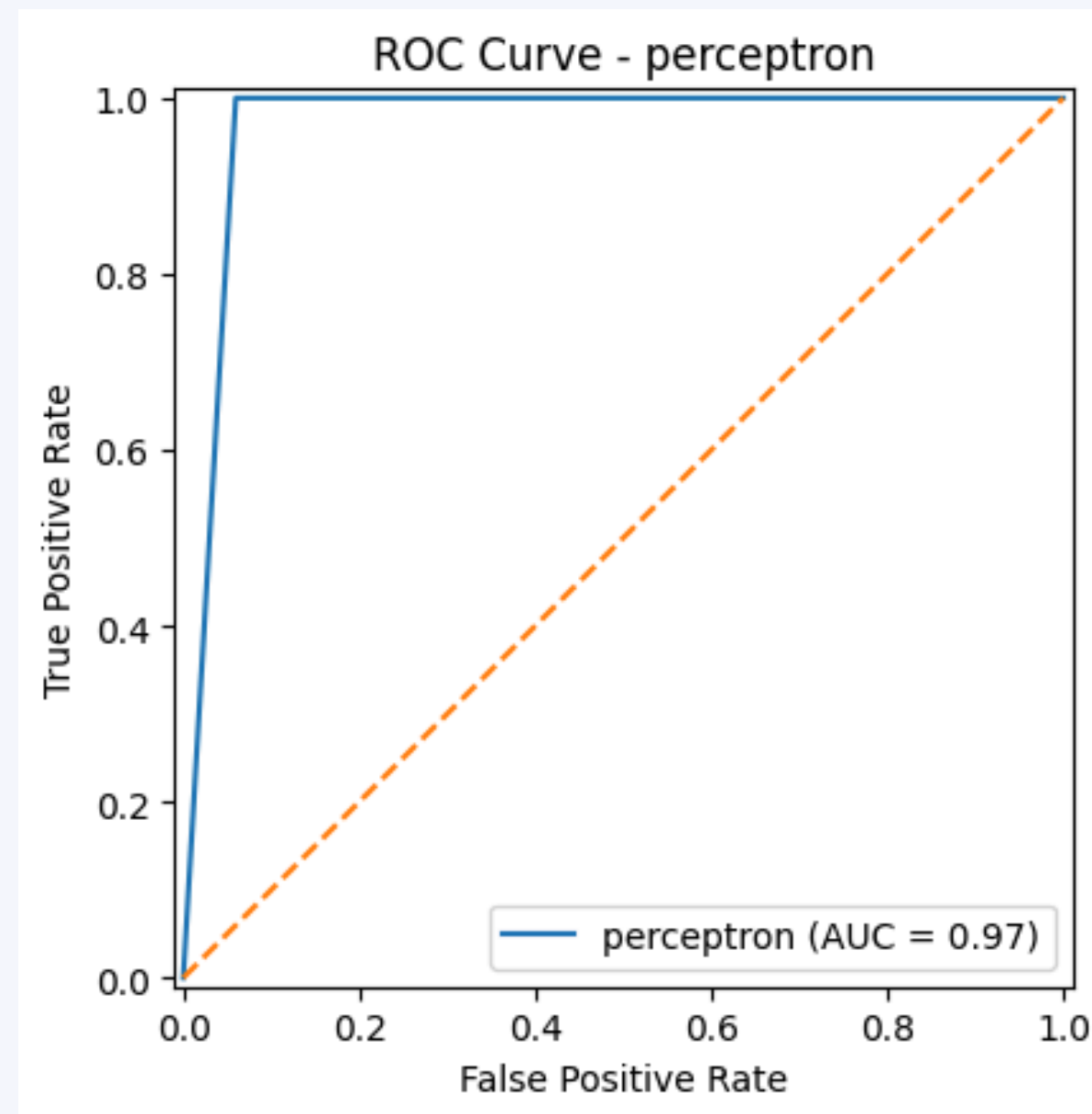


Modèle de perceptron

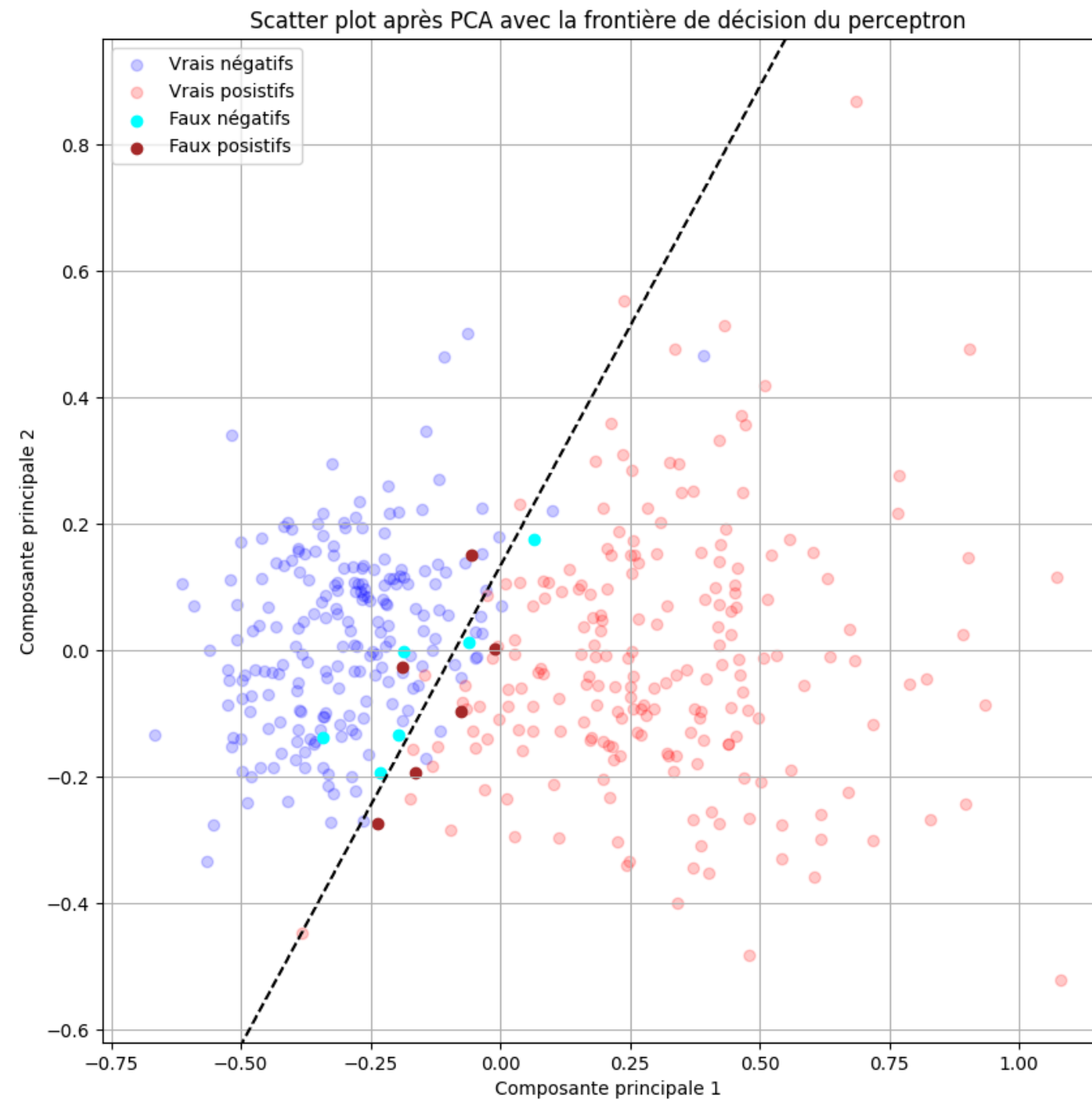


Résultat après backward features selction

Backward:
Accuracy : 96.88%
Precision : 100.0%
Recall : 93.85%
F1-Score : 96.83%

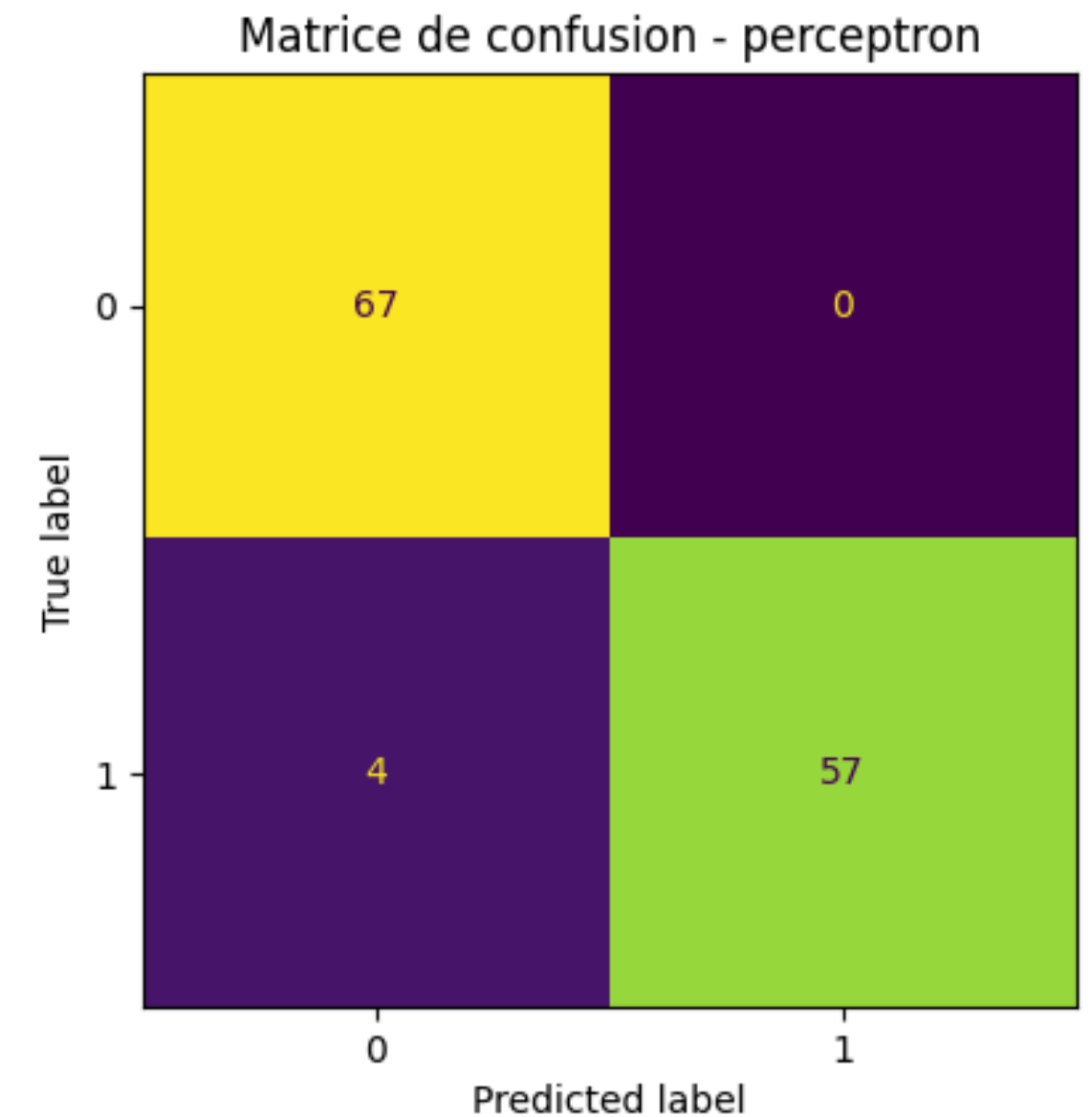
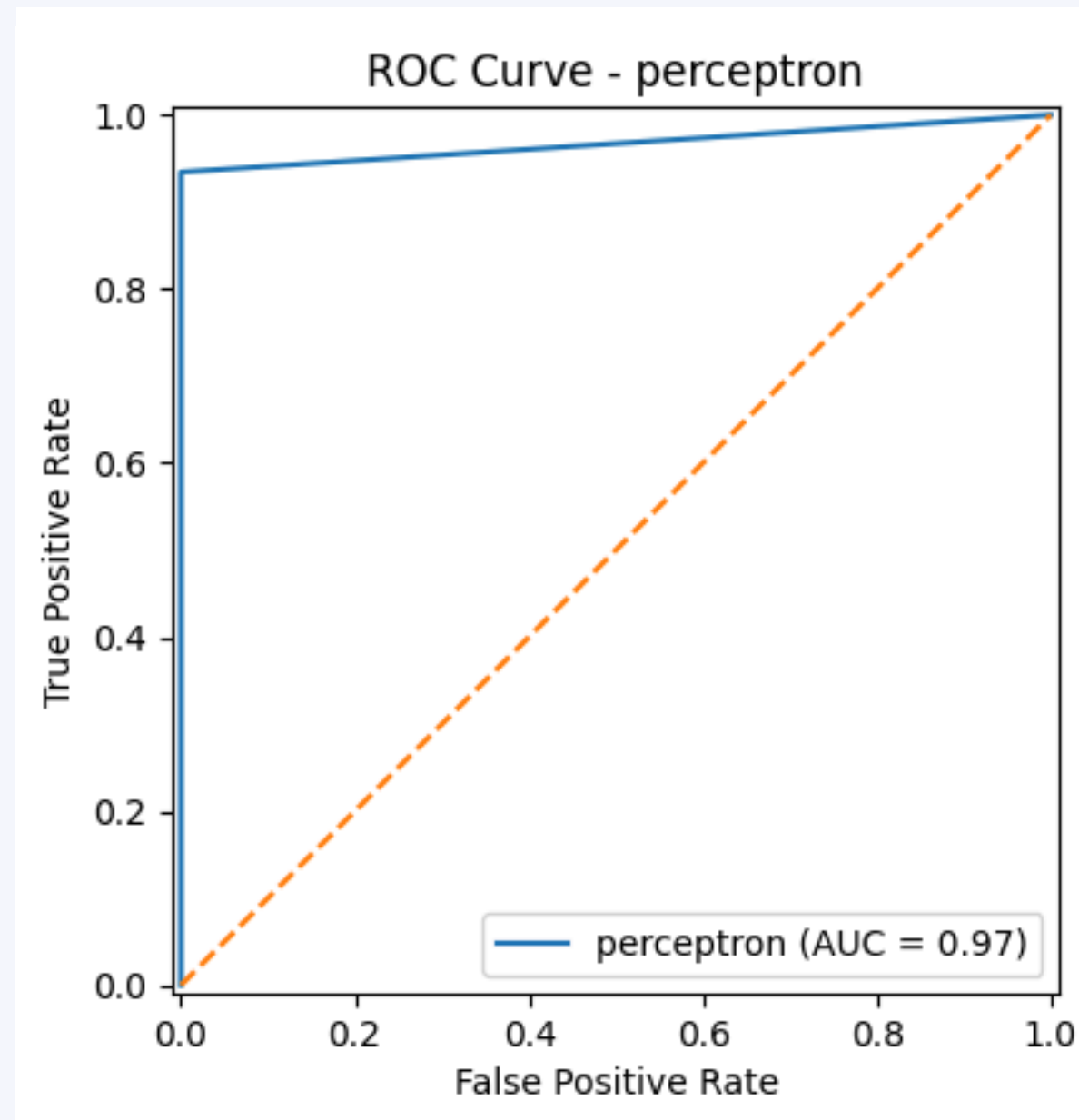


Résultat après backward features selction

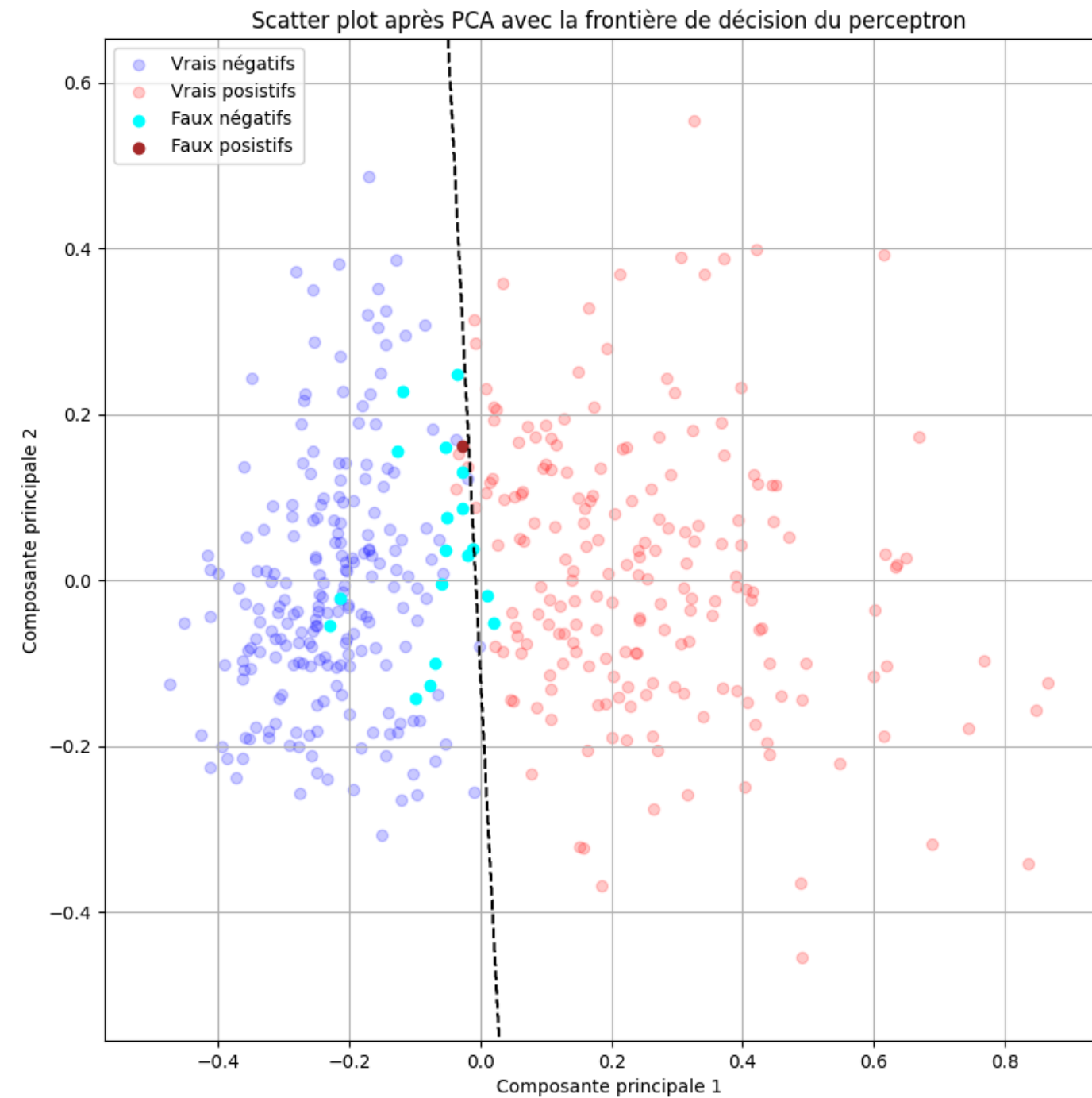


Résultats après forward features selection

Forward:
Accuracy : 96.88%
Precision : 93.44%
Recall : 100.0%
F1-Score : 96.61%



Résultats après forward features selection



Scores après application à la totalité des données

Backward features

Accuracy : 96.88%
Precision : 100.0%
Recall : 93.85%
F1-Score : 96.83%



Accuracy : 97.36%
Precision : 97.17%
Recall : 95.81%
F1-Score : 96.49%

Backward features

Accuracy : 96.88%
Precision : 93.44%
Recall : 100.0%
F1-Score : 96.61%



Accuracy : 96.66%
Precision : 91.51%
Recall : 99.49%
F1-Score : 95.33%



Piste d'amélioration

- Augmenter la taille des données.
- Améliorer leur qualité en évitant le data leak.
- Explorer la piste des réseaux de neurones pour acquérir un modèle plus complexe et précis

