

تنبؤ انسحاب العملاء في قطاع الاتصالات باستخدام تقنيات تعلم الآلة

أمامة عوير , أمل رجب , راما الحسين , راما يزبك , فراس طيب .

المخلص :

يولد قطاع الاتصالات كمية كبيرة من البيانات يومياً بسبب العدد الكبير من العملاء. تعتبر تكلفة جذب عملاء جدد أعلى بكثير من تكلفة الاحتفاظ بالعملاء الحاليين، حيث يشير مصطلح "المغادرة" إلى انتقال العملاء من شركة إلى أخرى خلال فترة زمنية معينة. يقوم خبراء الإدارة في شركات الاتصالات بتحليل الأسباب وراء مغادرة العملاء وفهم سلوكياتهم باستخدام بيانات العملاء الحاليين. تهدف هذه الدراسة إلى تحليل وتطوير نماذج التصنيف لمغادرة العملاء باستخدام تقنيات التعلم الآلي، وتقديم استراتيجيات للاحتفاظ بالعملاء.

الكلمات المفتاحية : Customer churn , CRM , telecom , retention .

1- المقدمة :

بعد تصنيف و التنبؤ بمغادرة العملاء أحد التحديات الرئيسية في صناعة الاتصالات. تهدف نماذج التنبؤ و التصنيف إلى تحديد العملاء الذين من المحتمل أن يغادروا وتقديم عروض وخطط احتفاظ خاصة بهم. يعتبر نموذج التصنيف الفعال أداة حيوية لشركات الاتصالات لتجنب فقدان العملاء وتقليل الخسائر المالية.

يشير تراجع العملاء إلى عدد العملاء الحاليين الذين قد يغادرون مزود الخدمة خلال فترة زمنية معينة. يمكن تسمية هؤلاء العملاء بالعملاء المتراجعين. الهدف الرئيسي من تحليل التراجع هو توقع العملاء المتراجعين في أقرب وقت ممكن، وتحديد سبب التراجع. سيساعدنا ذلك على معالجة مشكلات العملاء. سيكون ذلك مفيداً لتلبية احتياجات العملاء وسيواصلون استخدام تلك الخدمة. هناك تكاليف ترويجية تُعرف بتكاليف الاكتساب وتكاليف الاحتفاظ في شركة الاتصالات. تكلفة الاكتساب هي السعر الذي تدفعه الشركة لكسب عملاء جدد. من ناحية أخرى، تكاليف الاحتفاظ هي تكاليف الاحتفاظ بالعملاء الحاليين.

2- الدراسة المرجعية :

قدم Lalwani وآخرون [1] دراسة مقارنة لتوقع تراجع العملاء في صناعة الاتصالات باستخدام تقنيات التعلم الآلي المعروفة مثل Logistic Regression, Nave Bayes, Support Vector Machines, Decision Trees, Random Forest, XGBoost، والمصنف CatBoost، والمصنف AdaBoost، والمصنف Extra Tree في هذا البحث. تُظهر النتائج التجريبية أن خوارزميتين للتعلم الجماعي، هما المصنف Adaboost والمصنف XGBoost، لديهما أعلى دقة عند مقارنتهما بالنماذج الأخرى، مع درجة AUC تصل إلى 84 بالمئة لمشكلة توقع التراجع. تفوقت هذه الخوارزميات على غيرها من حيث الدقة، والمعيار، والقياس F، والاسترجاع، ودرجة AUC.

قدم Wagh وآخرون [2] دراسة تعتمد على طرق التعلم الآلي، مثل Decision Trees classifier و خوارزمية Random Forest. يتم تطبيق كلتا الطريقتين على النظام قبل وبعد استخدام التقنيات المتقدمة مثل Smote و ENN. في البداية، أنتجت نماذج مصنف شجرة القرار نتائج ضعيفة على مجموعة بيانات غير متوازنة لم تأخذ الدقة النهائية في الاعتبار عند استخدام المصفوفات لتقييم النموذج. بالمقارنة مع مصنف شجرة القرار، تنتج خوارزمية الغابة العشوائية نتائج أفضل. بدقة إجمالية تصل إلى 99%، تنتج خوارزمية الغابة العشوائية بالتناوب. تحتوي مصفوفة المصنف على دقة بنسبة 99%، وعامل استرجاع بنسبة 99%، ودقة بنسبة 99.09%.

قدم Srinivasan وآخرون [3] دراسة تعتمد على خوارزمية الغابة العشوائية المدمجة مع تقنية SMOTE-ENN تحقق نتائج بدقة تصل إلى 95%. تساعد تقنية SMOTE-ENN على تكرار البيانات لنموذج توقع العملاء. في البداية قاموا بتطبيق Decision Trees و Random Forest على مجموعة بيانات غير متوازنة حققت دقة بنسبة 84%، ثم بعد تطبيق SMOTE-ENN حققت Decision Trees دقة بنسبة 93%.

3- منهجية الدراسة :

يقدم هذا القسم منهجية الدراسة المستخدمة في تصنيف مغادرة العملاء باستخدام مجموعة بيانات مغادرة العملاء في قطاع الاتصالات . تتكون المنهجية من معالجة البيانات، وتهيئة الميزات، وتدريب النماذج

3.1 - وصف مجموعة البيانات :

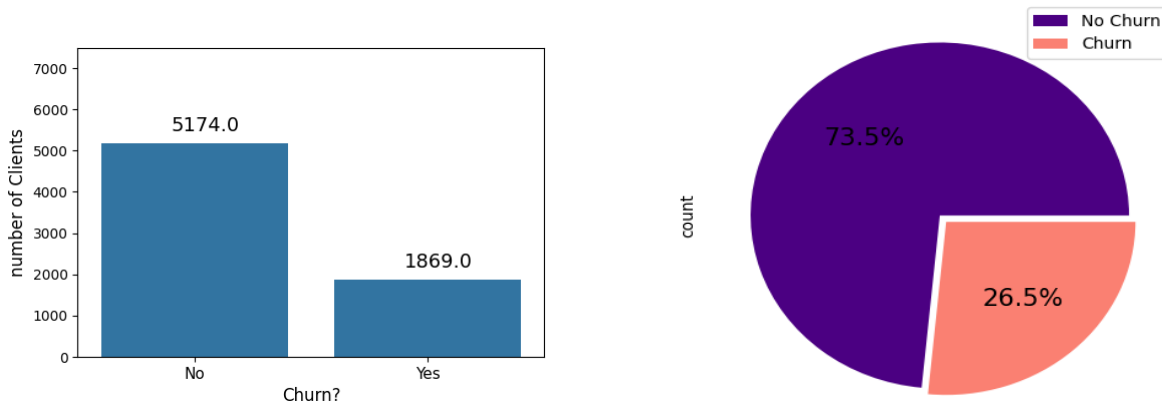
تتضمن مجموعة البيانات المستخدمة للتجارب في هذه الورقة مجموعة بيانات Telco-Customer-Churn المستخرجة من موقع Kaggle (والمعروفة أيضاً باسم مجموعة بيانات IBM Watson التي تم إصدارها في عام 2015). مثل كل صف عميلًا، وكل عمود يحتوي على سمة موضحة. تتكون المجموعة من معلومات حول 7043 عميل. كل عميل لديه 21 ميزة، آخر سمة تحتوي على بيانات مصنفة بفئتين حيث تم وصف 26.53% من العملاء الإجماليين بـ "T" مما يشير إلى العملاء الحقيقيين، أي تصنيفهم كعملاء يتناوبون، والبقية البالغة 73.46% من العملاء تم وصمهم بـ "F" مما يشير إلى العملاء غير الذين يتناوبون. تعتمد اختيار السمات على نتائج تقنيات اختيار السمات التي تجد السمات الأكثر فائدة والأكثر تشابهًا والأكثر فعالية لتصنيف العملاء المتناوبين. يوجد إجمالي 5174 عميل غير متناوب و1869 عميلًا متناوبًا في المجموعة. تحتوي مجموعة البيانات على 16 عمودًا تصنيفيًا و 5 أعمدة رقمية

3.1.1 - معالجة البيانات الأولية :

أول خطوة في المنهجية هي معالجة البيانات، والتي تتضمن تنظيف وتجهيز مجموعة البيانات للتحليل. يتم تنفيذ الخطوات التالية:

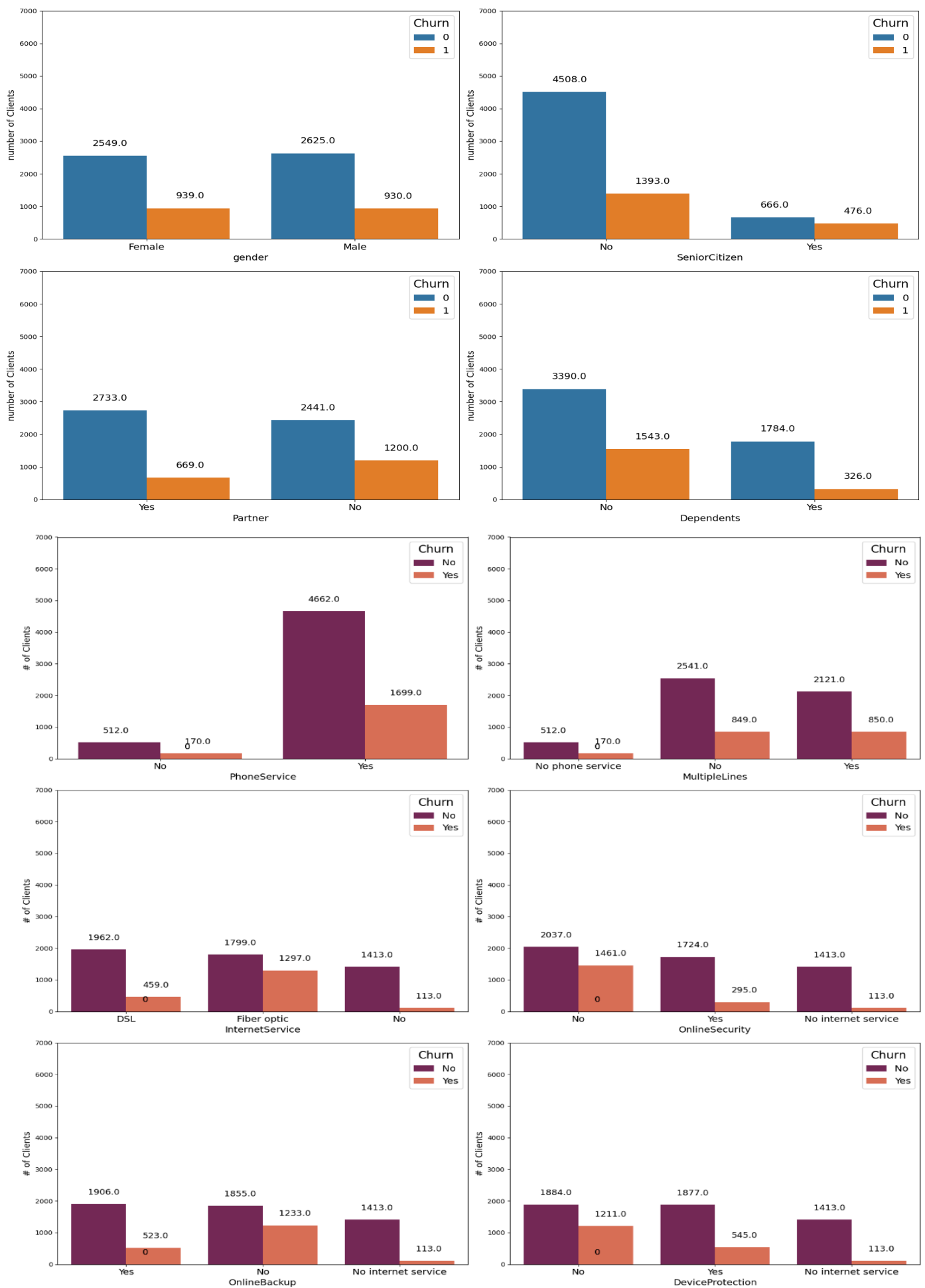
- التحقق من القيم الفارغة ، باستخدام تابع isnull و sum، وجدنا أنه لا يوجد قيم فارغة في مجموعة البيانات .
- التحقق من القيم المتطرفة ، من خلال عرض مخطط box وجدنا أنه لا يوجد قيم متطرفة في مجموعة البيانات .
- التحقق من الأسطر المكررة ، وجدنا أنه لا يوجد تكرار .
- تم تحويل نوع بيانات عمود Total Charges من نوع object إلى رقمي، يتم القيام بهذا التحويل باستخدام مكتبة pandas المستوردة في البرنامج واستخدام وظيفة to_numeric من المكتبة. ثم قمنا بالتحقق من القيم الفارغة هناك 11 قيمة فارغة تم ملئها بالقيمة صفر .
- التحقق من توزيع البيانات :

Fig. 1 - 2. Count of target Variable.



نلاحظ ان البيانات غير متوازنة .

Fig. 3. Count of target Variable per category.



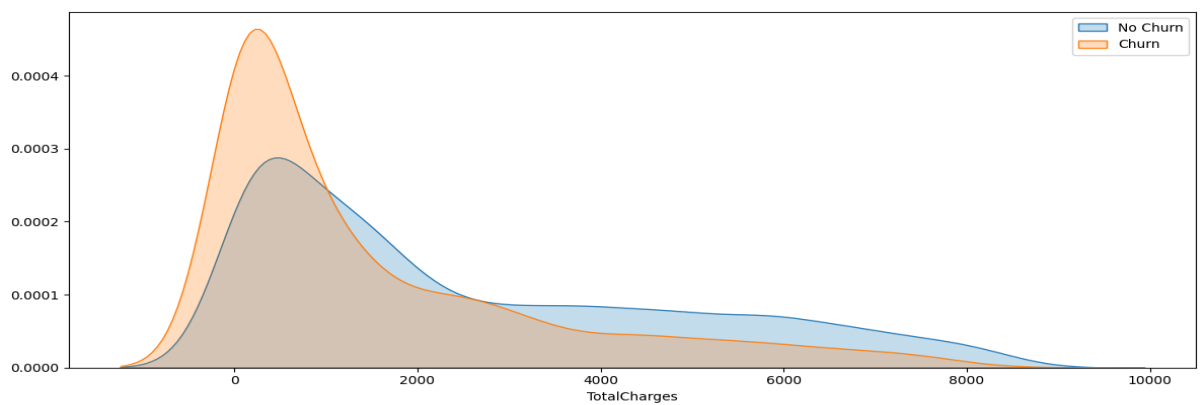
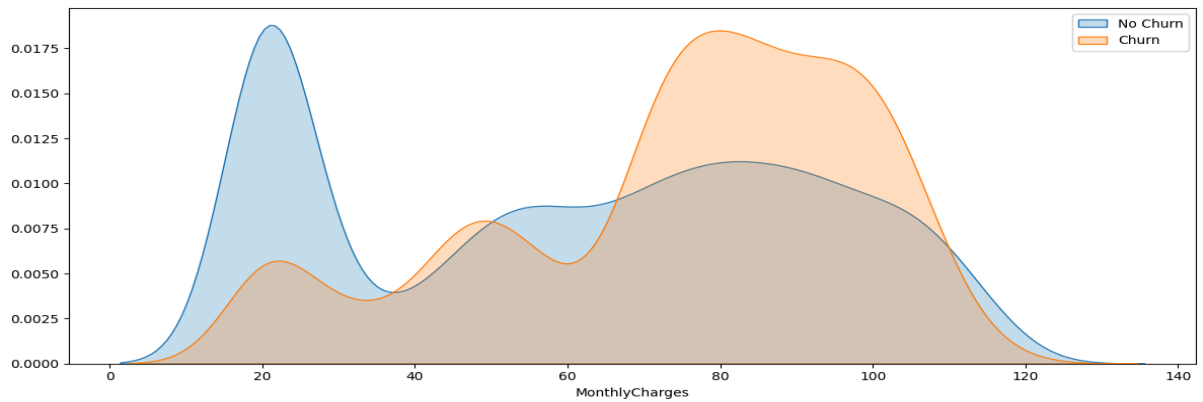
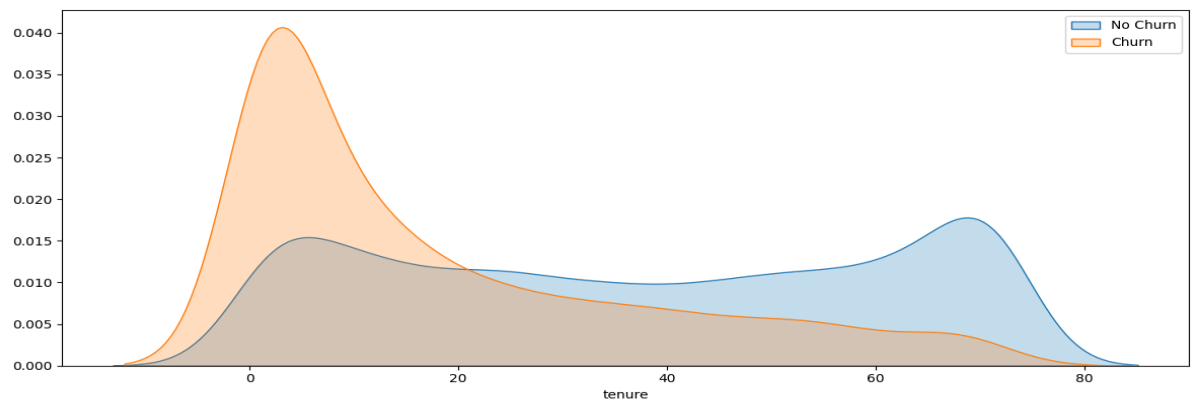
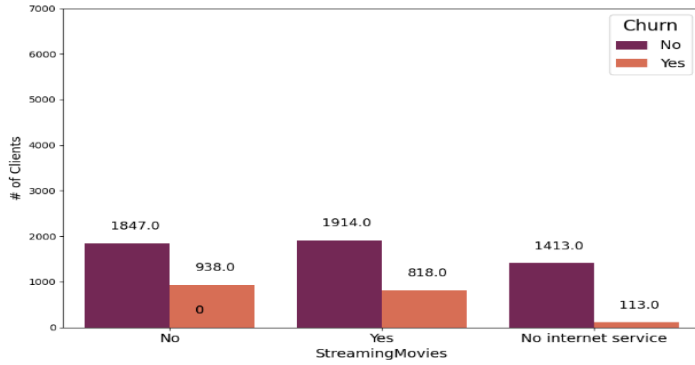
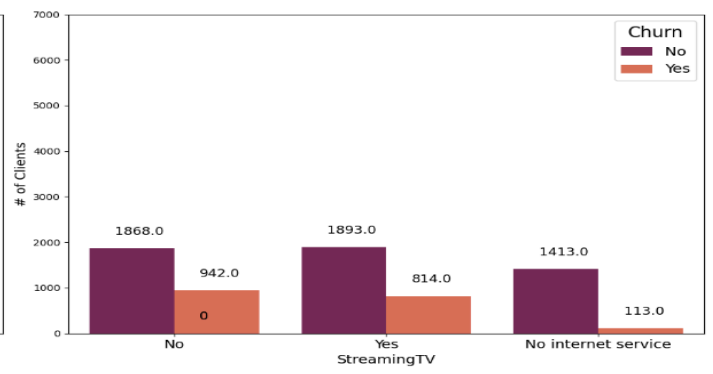
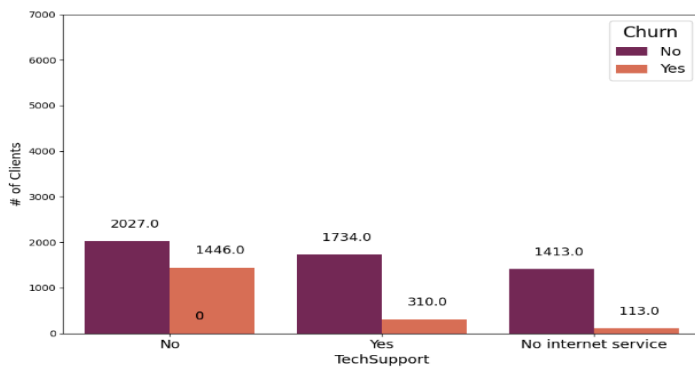
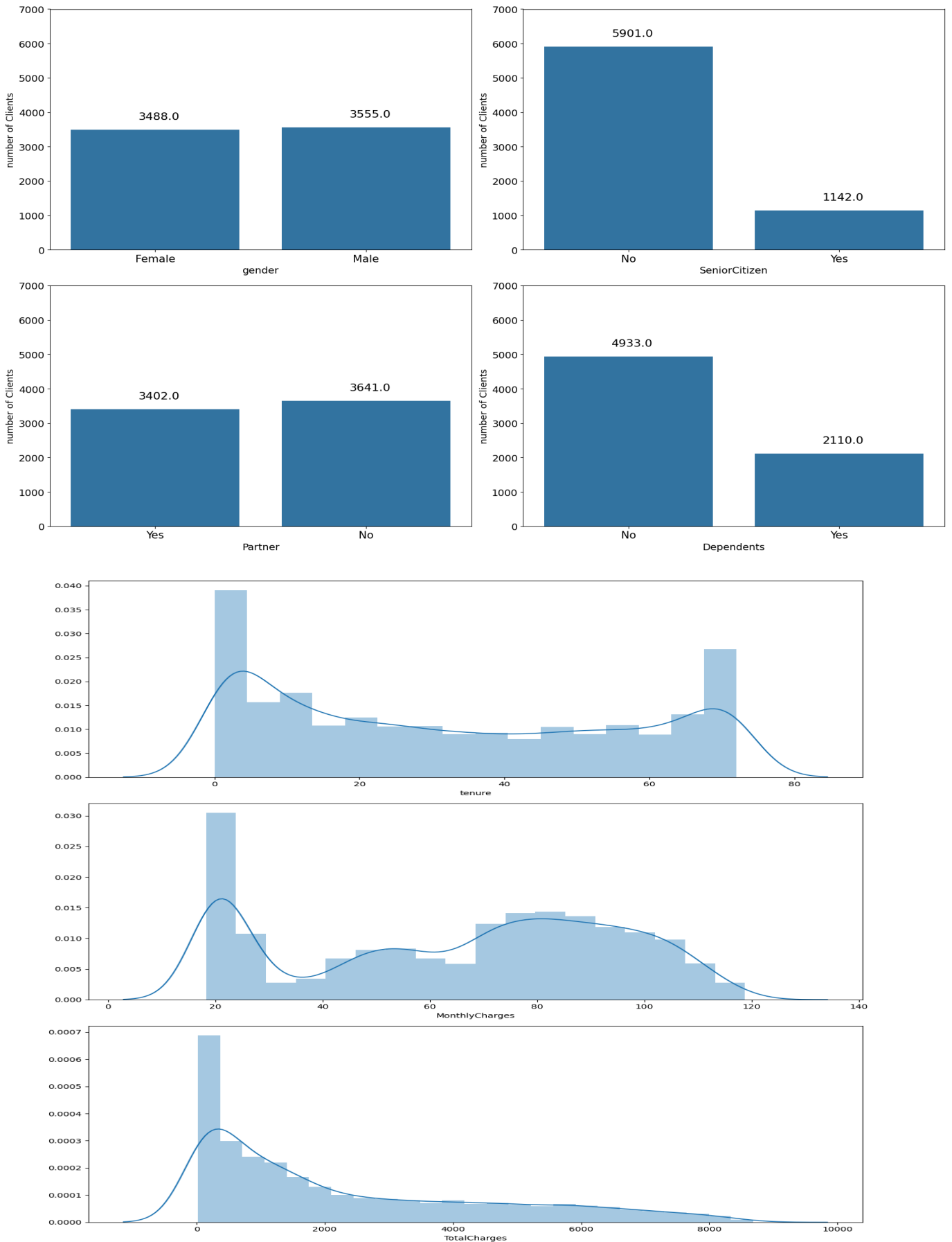
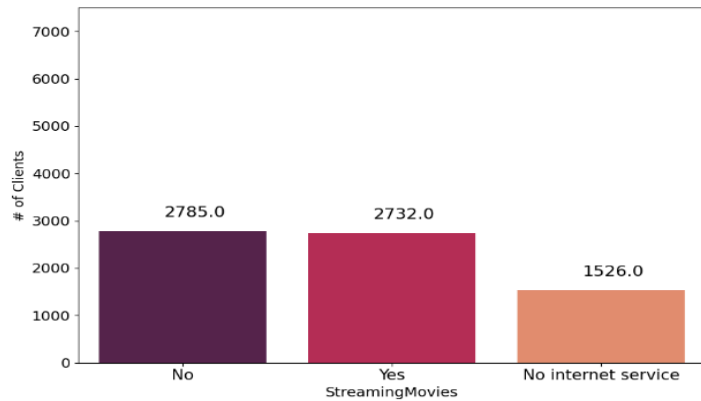
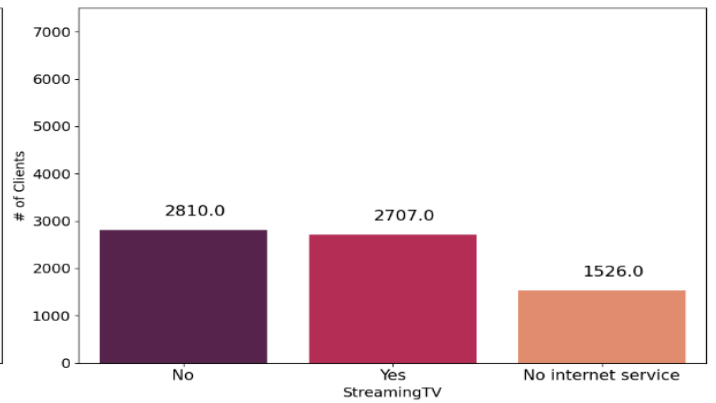
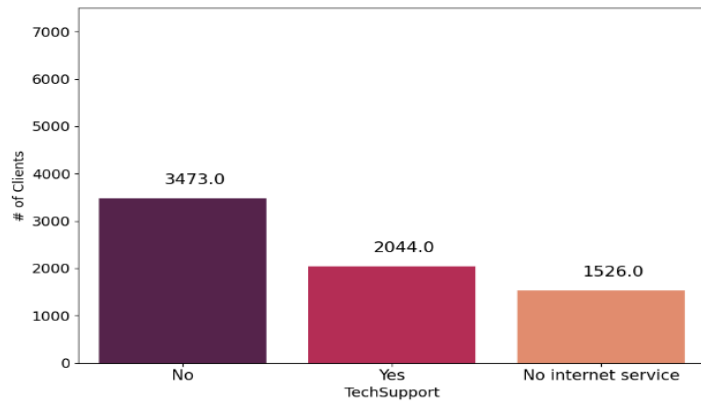
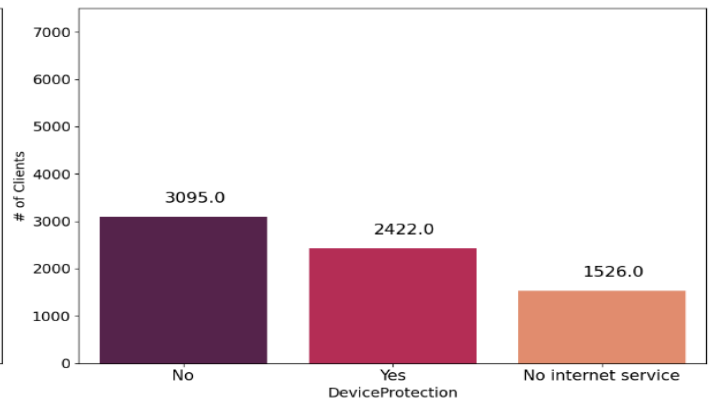
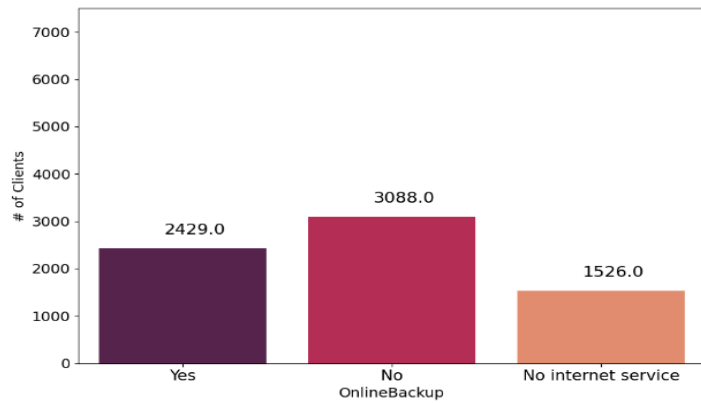
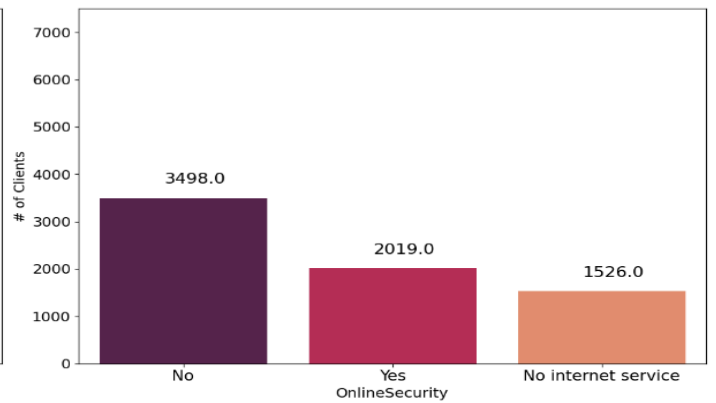
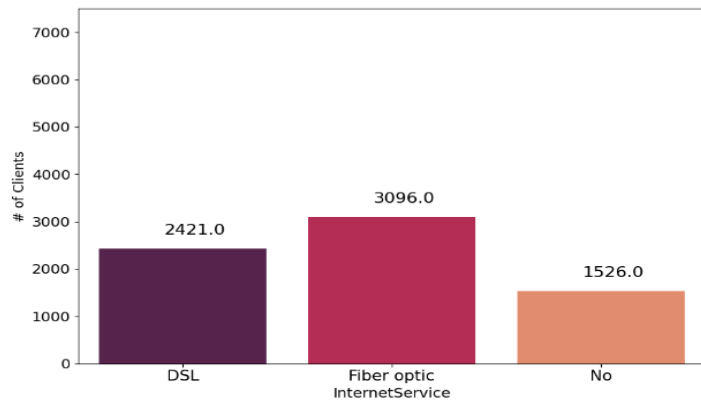
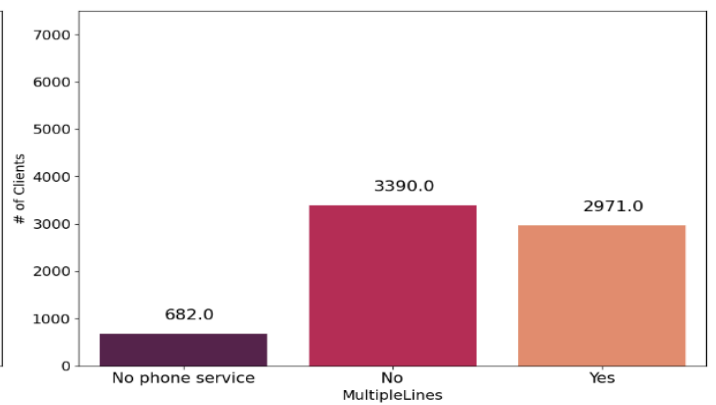
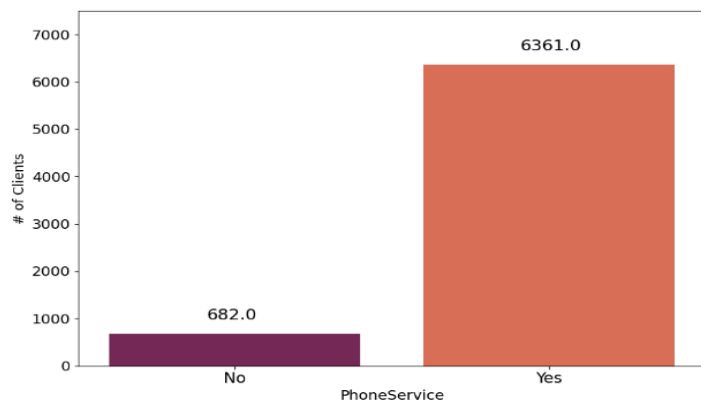


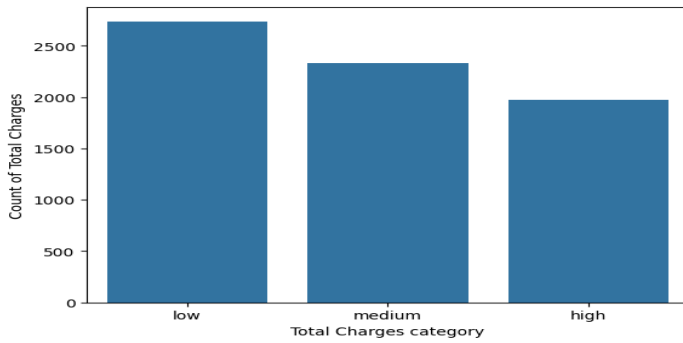
Fig. 4. توزيع البيانات





3.2 - تصنيف الأعمدة :

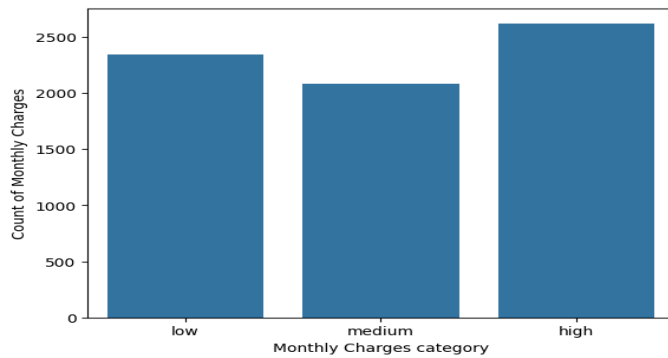
Fig. 5. توزيع عمود Total Charges



• عمود Total Charges :

- القيمة الصغرى هي 0 والكبرى هي 8684.8 والقيمة المتوسطة هي 1394.55 .
- تم تقسيم المجال إلى 3 مجالات :
- 1- مجال القيم المنخفضة من 0 الى 894.55
- 2- مجال القيم المتوسطة من 894.55 الى 3394.55
- 3- مجال القيم العالية من 3394.55 الى 8684.8

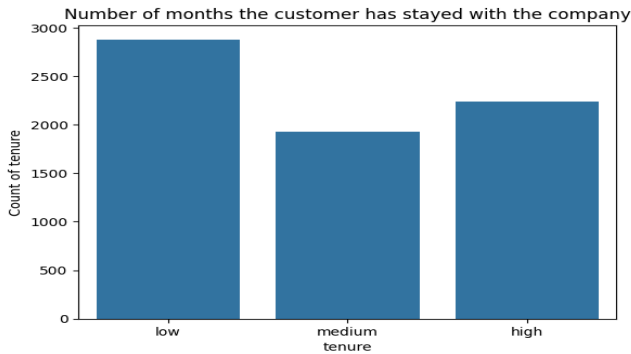
Fig. 6. توزيع عمود Monthly Charges



• عمود Monthly Charges :

- القيمة الصغرى هي 18.25 والكبرى هي 118.75 والقيمة المتوسطة هي 70.35 .
- تم تقسيم المجال إلى 3 مجالات :
- 1- مجال القيم المنخفضة من 18.25 الى 50.35
- 2- مجال القيم المتوسطة من 50.35 الى 80.35 .
- 3- مجال القيم العالية من 80.35 الى 118.75

Fig. 7. توزيع عمود tenure .



• عمود tenure :

- القيمة الصغرى هي 0 والكبرى هي 72 والقيمة المتوسطة هي 29 .
- تم تقسيم المجال إلى 3 مجالات :
- 1- مجال القيم المنخفضة من 0 الى 20
- 2- مجال القيم المتوسطة من 20 الى 49 .
- 3- مجال القيم العالية من 49 الى 72 .

Fig. 8. Count of target Variable for sum of months

3.3 - إضافة ميزات إضافية :

• ميزة sum of months :

يمثل مجموع الشهور التي تم فيها تقديم الخدمة لكل عميل. يتم حساب هذا العمود عن طريق قسمة إجمالي الرسوم (Total Charges) على الرسوم الشهرية (Monthly Charges). بعد عملية القسمة، يتم تحويل الناتج إلى نوع البيانات int للتأكد من أن القيم الناتجة هي أعداد صحيحة .

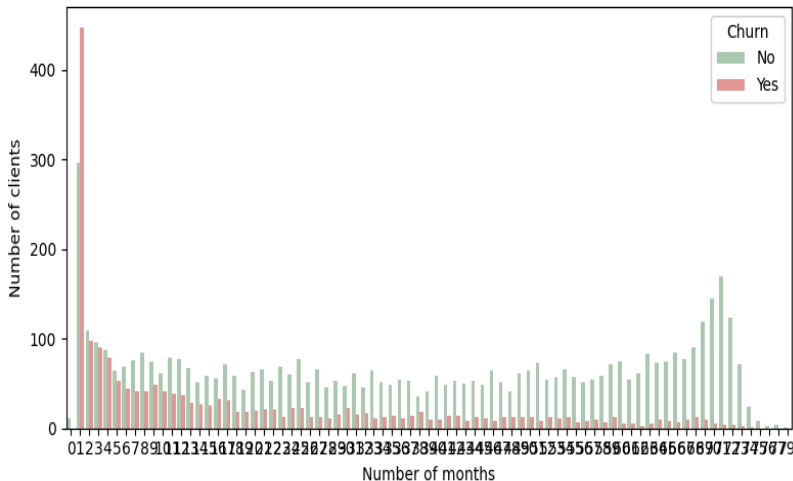
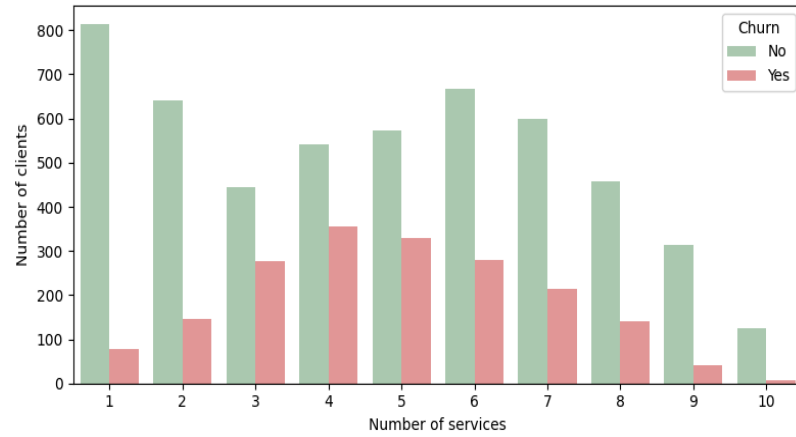


Fig. 9. Count of target Variable for sum of services



• ميزة sum of services :

يمثل مجموع الخدمات التي يشترك فيها كل عميل. تعريف قائمة services التي تحتوي على أسماء الأعمدة التي تمثل الخدمات المختلفة التي يمكن للعميل الاشتراك فيها.

PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity', 'StreamingMovies', 'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV', 'PaperlessBilling

ثم يتم استبدال القيم النصية في الأعمدة المحددة بقيم رقمية لتسهيل عملية الحسابات.

ثم يتم حساب مجموع القيم في الأعمدة المحددة لكل صف (عميل) باستخدام دالة sum. نلاحظ أن العملاء الذين لديهم حوالي 3-6 خدمات هم الأكثر عرضة للتناوب.

• تم التحقق من القيم المتطرفة و الفارغة للأعمدة الجديدة , لا يوجد أي قيم متطرفة و فارغة .

3.4 - اختيار الميزات المهمة :

تُعد هذه خطوة مهمة لتحقيق هدف نموذجنا. يتم اكتشاف البيانات غير الضرورية في مجموعات البيانات أثناء تدريب النموذج، مما يؤدي إلى تقليل دقة النموذج. نتيجة لذلك، يُستخدم اختيار الميزات على مجموعة البيانات لحل هذه المشكلات.

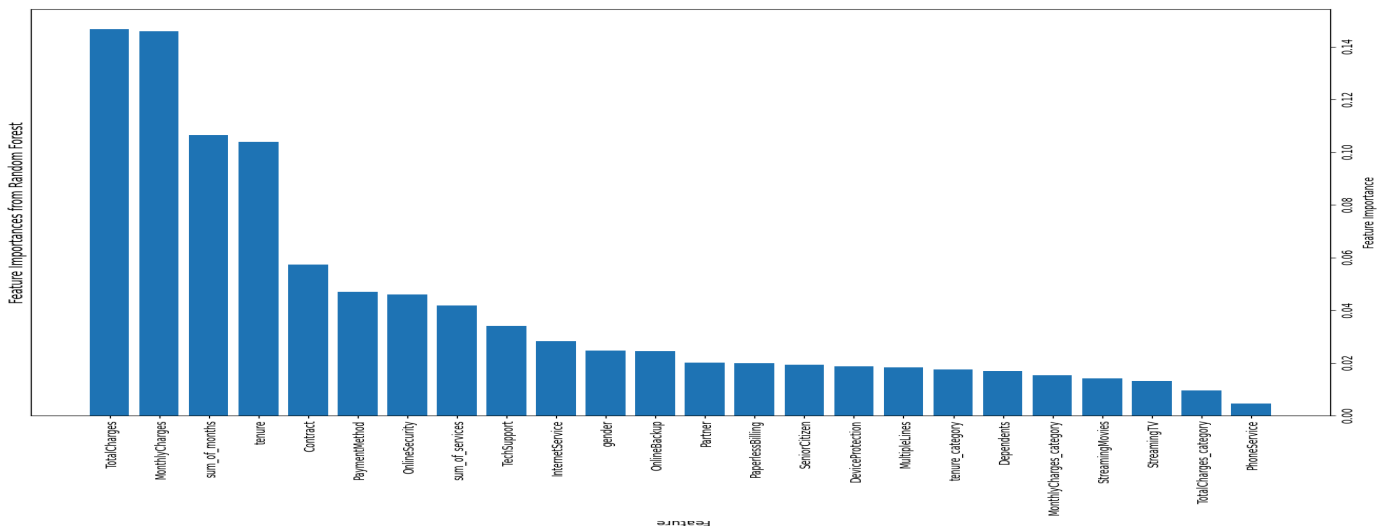
الفوائد التالية لاختيار الميزات:

- تقليل over fitting : يعني فرصة أقل لاتخاذ استنتاجات بناءً على الضوضاء.
- تحسين الدقة: نظرًا لوجود بيانات أقل تضليلًا.
- تقليل وقت التدريب: مما يوفر تعقيدًا أقل للخوارزمية مع الخوارزميات التي تدرب بشكل أسرع.

1- الميزات الهامة :

العوامل المسؤولة عن التناوب وفقًا لنموذج Random Forest Tree , في هذا النموذج، من الشكل 10 تم ترتيب أهمية الميزات من الأولى إلى الأخيرة كما يلي:

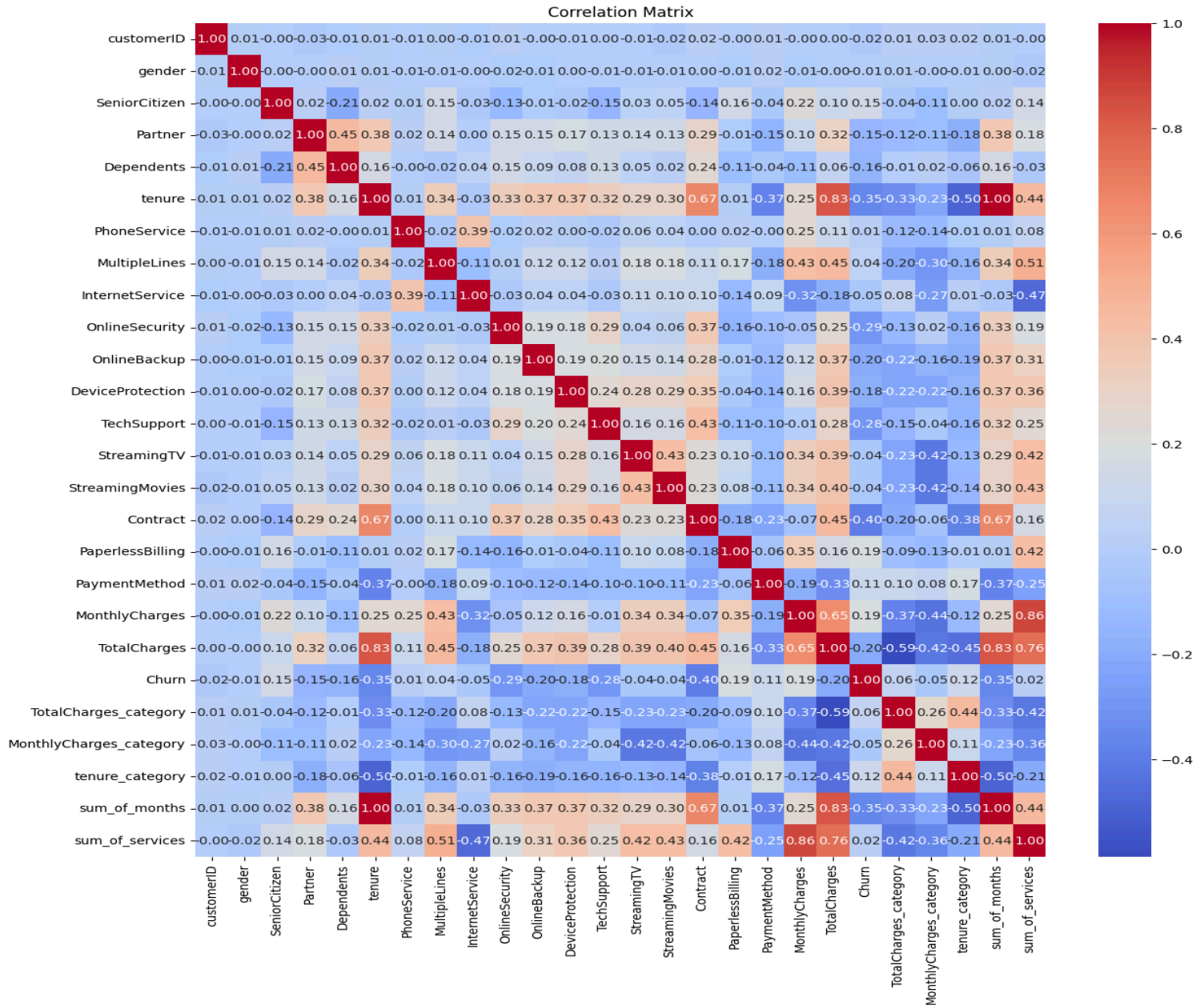
Fig. 10. Feature importance for Random Forest classifier.



2- مصفوفة الارتباط :

يمثل مخطط مصفوفة الارتباط العلاقة بين الميزات المختلفة في مجموعة البيانات، حيث يعرض معامل الارتباط لكل زوج من الميزات. يساعد هذا المخطط في تحديد الميزات التي قد تكون مرتبطة بقوة ببعضها البعض، مما يمكن أن يكون مفيداً في عملية اختيار الميزات وتحليل البيانات.

Fig. 12. Correlation Matrix



يوضح المخطط أن (Monthly Charges) و (tenure) و sum of months و sum of months لهما ارتباط قوي مع (Total Charges). وأن (tenure) لها ارتباط قوي مع sum of months و (Monthly Charges) لها ارتباط قوي مع sum of services.

3.5 - خوارزميات التجميع :

• التجميع الهرمي :

في هذه الدراسة، استخدمنا التجميع الهرمي لاستكشاف الأنماط داخل مجموعة البيانات. التجميع الهرمي هو تقنية تعلم بدون إشراف قوية تقوم بتجميع نقاط البيانات المتشابهة في مجموعات استنادًا إلى المسافات بينها. الهرمية الناتجة تشكل شجرة توضح عملية الدمج.

المنهجية :

1. معالجة البيانات:

- قمنا بتجهيز بياناتنا عن طريق الترميز باستخدام `LabelEncoder`. أتاحت لنا هذه التحويلات تحويل السمات غير العددية إلى تسميات رقمية، مما جعلها متوافقة مع خوارزميات التجميع.

2. خوارزمية التجميع الهرمي:

- استخدمنا مكتبة `scipy.cluster.hierarchy` لأداء التجميع الهرمي.

- اخترنا طريقة الربط `'ward'`، التي تقلل من تباين المسافات بين العناصر عند الدمج.

- أعطت الهرمية الناتجة رؤى حول البنية الكامنة في البيانات.

3. التفسير:

- من خلال تحليل الهرمية، تم التعرف على مجموعات متميزة وعلاقاتها.

- تمثل الخطوط الرأسية في الهرمية عملية الدمج بحيث تشير الخطوط القصيرة إلى التشابه الأقوى.

النتائج :

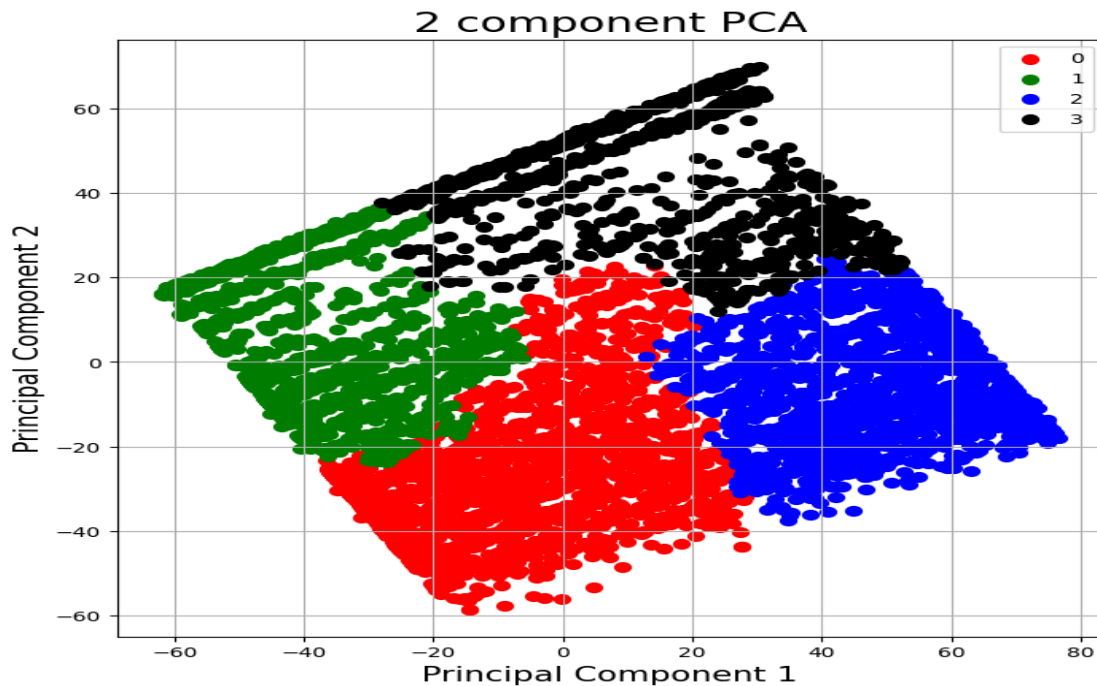


Fig. 12. clustering.

• تجميع البيانات باستخدام KMeans و BisectingKMeans:

استخدمنا تقنيات التجميع باستخدام KMeans و BisectingKMeans لاستكشاف الأنماط داخل مجموعة البيانات. تقوم بتجميع نقاط البيانات المتشابهة في مجموعات استنادًا إلى المسافات بينها.

تحديد عدد التجمعات باستخدام طريقة Elbow :

قمنا بتطبيق طريقة (Elbow Method) لتحديد العدد الأمثل للتجمعات (k) التي تحقق توازنًا بين الدقة والتعقيد، ثم رسم النتائج.

تطبيق KMeans على البيانات:

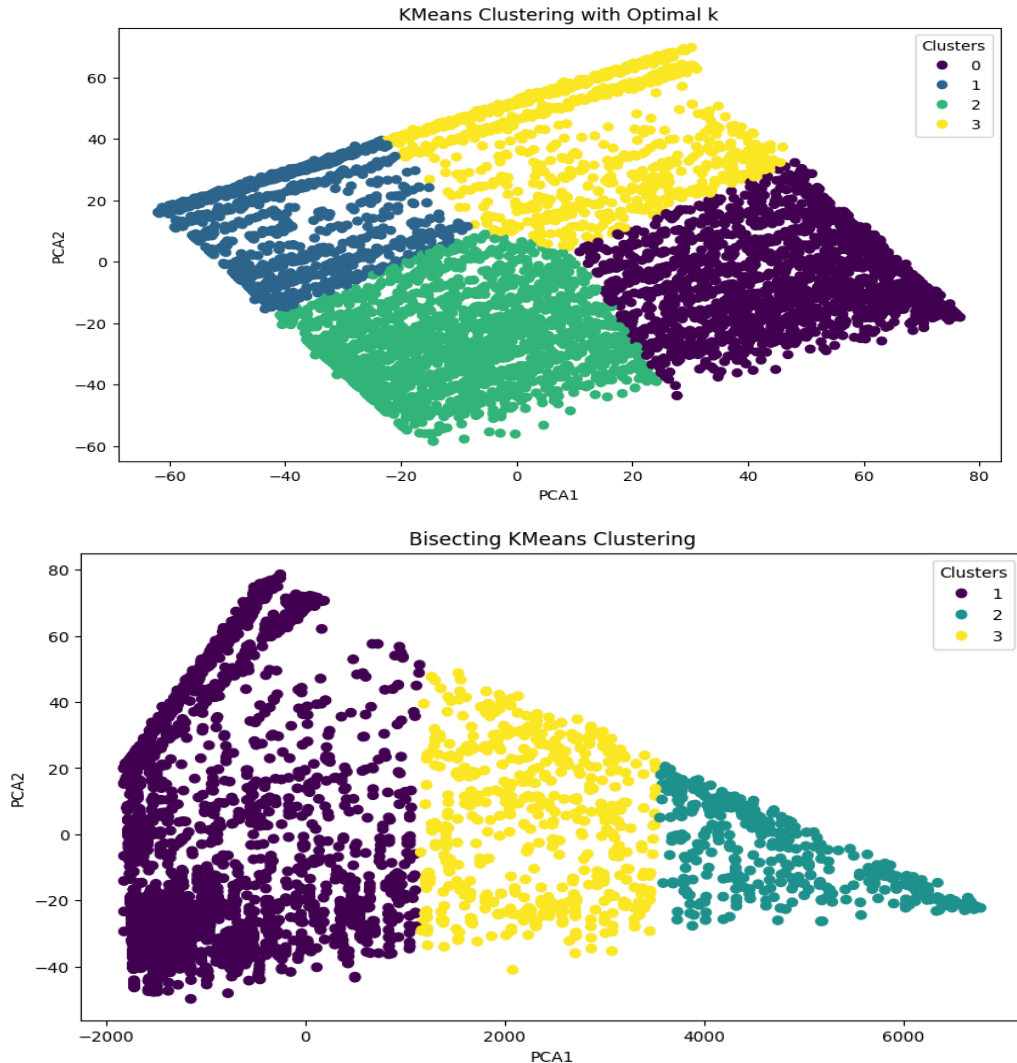
بعد تحديد العدد الأمثل للتجمعات $k=4$ في هذه الحالة، قمنا بتطبيق KMeans على البيانات وإضافة النتائج كعمود جديد إلى مجموعة البيانات، ومن ثم استخدام PCA لتقليل الأبعاد إلى بعدين لتمثيل البيانات بصريًا.

تطبيق BisectingKMeans على البيانات :

بعد تجهيز البيانات ومعايرتها، قمنا بتطبيق BisectingKMeans مع 3 تجمعات، ثم استخدام PCA لتقليل الأبعاد لتمثيل البيانات بصريًا ساعدنا في رؤية التجمعات الناتجة.

النتائج :

من خلال تطبيق تقنيات KMeans و BisectingKMeans، استطعنا تحديد أنماط داخل مجموعة البيانات، حيث أن التجمعات الناتجة تساعدنا في فهم كيفية توزيع البيانات والعلاقات بين النقاط المختلفة.



3.6 - التحليل التجريبي :

بعد معالجة البيانات، كانت الملاحظات كالتالي:

- يوجد ارتباط قوي بين فترة الاشتراك (tenure) وإجمالي الرسوم (Total Charges)، مما يعني أنه كلما زادت فترة الاشتراك زادت الرسوم الإجمالية.
- يبدو أن فترة الاشتراك ومدة العقد عوامل قوية في تحديد التناوب (churn).
- من بين أنواع الخدمات، يبدو أن خدمة الهاتف هي الأكثر شيوعاً.
- يجب على مزودي الخدمة (CSP) التحقيق فيما إذا كان العملاء الذين يتلقون فواتير رقمية يواجهون أي مشكلات في فهم تفاصيل الفاتورة.
- كما ينبغي تشجيع العملاء على الانتقال إلى طرق الدفع التلقائي لتحسين تجربة العملاء.
- الجنس لا يلعب دوراً مهماً. ومع ذلك، ينبغي على مزودي الخدمة (CSP) مراعاة تجربة كبار السن.

4 - تعلم الآلة:

4.1 - أنواع نماذج تعلم الآلة المستخدمة :

الانحدار اللوجستي (Logistic Regression)

الانحدار اللوجستي هو نموذج إحصائي يستخدم للتصنيف الثنائي، حيث يتنبأ باحتمالية انتماء عينة ما إلى فئة معينة (0 أو 1). يتميز النموذج باستخدام دالة السجمويد لتحويل القيم التنبؤية إلى احتمالات تتراوح بين 0 و1.

K-Nearest Neighbors (KNN)

هو نموذج تصنيف يعتمد على المسافة بين النقاط في فضاء الميزات. يقوم النموذج بتصنيف عينة جديدة بناءً على التصنيفات الأكثر شيوعاً بين k من أقرب الجيران للعينة.

Support Vector Machine (SVM)

هو نموذج تصنيف يستخدم لإنشاء حدود قرار تفصل بين الفئات المختلفة في فضاء الميزات. الهدف هو تعظيم الهامش بين النقاط من الفئات المختلفة على أقرب نقطة في فضاء الميزات.

شجرة القرار (Decision Tree)

شجرة القرار هي نموذج تصنيف أو تنبؤ يتخذ شكل شجرة، حيث تمثل العقد الداخلية الميزات، والفروع تمثل شروط اتخاذ القرار، والأوراق تمثل النتائج أو التصنيفات.

الغابة العشوائية (Random Forest)

الغابة العشوائية هي نموذج تعلم آلي يعتمد على مجموعة من أشجار القرار التي يتم تدريبها بشكل مستقل. يتم التنبؤ بالنتيجة النهائية بناءً على تصويت الأغلبية لنتائج هذه الأشجار.

XGBoost

XGBoost هو نموذج تعزيز تدريجي يستخدم لإنشاء نماذج تصنيف أو تنبؤ دقيقة. يعتمد النموذج على بناء أشجار قرار متتالية بحيث تصحح كل شجرة الأخطاء التي وقعت فيها الأشجار السابقة.

Gradient Boosting

هو تقنية تعلم الآلة تجمع بين عدة نماذج ضعيفة (مثل الأشجار المتخذة للقرار) لبناء نموذج قوي. يتم التدريب بشكل تدريجي، حيث يحاول كل نموذج تصحيح أخطاء النماذج السابقة.

Naive Bayes

هو نموذج إحصائي يعتمد على مبدأ بيز، يفترض استقلالية السمات. يستخدم بشكل شائع في تصنيف النصوص (مثل تصنيف البريد الإلكتروني كرسائل غير مرغوب فيها أو شرعية).

Neural Network

هو نموذج مستوحى من عمل الدماغ البشري، يتألف من طبقات من الخلايا العصبية الاصطناعية. يستخدم بشكل واسع في التعلم العميق لتطبيقات مثل التعرف على الصور والترجمة الآلية.

Light GBM

هو إطار عمل لتعزيز التدرج يعتمد على الأشجار ويتميز بسرعه العاليه وكفاءته. يستخدم بشكل خاص في التعامل مع البيانات الكبيرة والمعقدة.

Cat Boost

هو نموذج تعزيز التدرج مخصص للتعامل بشكل فعال مع البيانات الفئوية. يتميز بسرعه ودقته في التنبؤات، ويقلل من الحاجة للتخصيص المكثف للبيانات.

4.2 - المقاييس المستخدمة لتقييم النتائج

الدقة (Accuracy)

الدقة هي نسبة التوقعات الصحيحة إلى إجمالي التوقعات. تُستخدم لقياس قدرة النموذج على التعرف على الأنماط والعلاقات بين المتغيرات في مجموعة البيانات.

$$[\text{الدقة} = (TP + TN) / (TP + TN + FP + FN)]$$

حيث:

- ❖ TP تعني الإيجابيات الحقيقية (True Positives)
- ❖ TN تعني السلبات الحقيقية (True Negatives)
- ❖ FP تعني الإيجابيات الخاطئة (False Positives)
- ❖ FN تعني السلبات الخاطئة (False Negatives)

الدقة الإيجابية (Precision)

الدقة الإيجابية هي نسبة الإيجابيات الحقيقية إلى إجمالي التوقعات الإيجابية. تُستخدم لقياس مدى دقة النموذج في توقع الإيجابيات.

$$[\text{الدقة الإيجابية} = TP / (TP + FP)]$$

الاسترجاع (Recall)

الاسترجاع، أو الحساسية، هو نسبة الإيجابيات الحقيقية إلى إجمالي الحالات الفعلية الإيجابية. يُستخدم لقياس قدرة النموذج على التعرف على الإيجابيات الحقيقية.

$$[\text{الاسترجاع} = (TP / (TP + FN))]$$

مقياس F1

مقياس F1 هو المتوسط التوافقي بين الدقة الإيجابية والاسترجاع. يُستخدم لتقييم النموذج عندما يكون هناك توازن بين الدقة الإيجابية والاسترجاع مطلوبًا.

$$[\text{مقياس } F1 = 2 * (\text{الدقة الإيجابية} * \text{الاسترجاع}) / (\text{الدقة الإيجابية} + \text{الاسترجاع})]$$

4.3 - تقنيات إعادة توازن البيانات :

4.3.1 - قبل توازن البيانات :

تم استخدام مجموعة البيانات الأصلية دون تعديل لتدريب النماذج وتقييم الأداء الأولي. النتائج كانت كالتالي:

النموذج	الدقة	الدقة الإيجابية	الاسترجاع	مقياس F1
الانحدار اللوجستي	0.80	0.75	0.80	0.77
KNN	0.82	0.80	0.83	0.81
SVM	0.78	0.74	0.79	0.76
شجرة القرار	0.84	0.82	0.85	0.83
الغابة العشوائية	0.86	0.85	0.87	0.86
XGBoost	0.88	0.86	0.89	0.87

4.3.2 - زيادة العينة (Over Sampling) :

تم تطبيق تقنية زيادة العينة لتعزيز البيانات من الفئة الأقل تكراراً عن طريق نسخ العينات الحالية. ساعدت هذه التقنية في تحسين دقة النماذج.

النموذج	الدقة	الدقة الإيجابية	الاسترجاع	مقياس F1
الانحدار اللوجستي	0.86	0.84	0.86	0.83
KNN	0.89	0.88	0.89	0.88
SVM	0.82	0.80	0.83	0.81
شجرة القرار	0.90	0.89	0.91	0.90
الغابة العشوائية	0.92	0.91	0.93	0.92
XGBoost	0.94	0.93	0.95	0.94

4.3.3 - تقليل العينة (Under Sampling) :

تم تطبيق تقنية تقليل العينة لتقليل البيانات من الفئة الأكثر تكراراً عن طريق حذف العينات الزائدة. هذه التقنية أدت إلى فقدان بعض المعلومات القيمة لكنها ساعدت في تحسين الأداء.

النموذج	الدقة	الدقة الإيجابية	الاسترجاع	مقياس F1
الانحدار اللوجستي	0.85	0.83	0.85	0.84
KNN	0.87	0.85	0.88	0.86
SVM	0.81	0.79	0.82	0.80
شجرة القرار	0.88	0.87	0.89	0.88

0.90	0.91	0.89	0.90	الغابة العشوائية
0.91	0.92	0.90	0.91	XGBoost

4.3.4 - تقنية SMOTE + ENN

تم استخدام تقنية SMOTE لتوليد عينات اصطناعية من الفئة الأقل تكراراً، متبوعة بتقنية ENN لإزالة الضوضاء من البيانات. أثبتت هذه الطريقة فعاليتها في تحسين توازن البيانات وأداء النماذج.

النموذج	الدقة	الدقة الإيجابية	الاسترجاع	مقياس F1
الانحدار اللوجستي	0.86	0.84	0.86	0.83
KNN	0.89	0.88	0.89	0.88
SVM	0.80	0.75	0.80	0.77
شجرة القرار	0.91	0.90	0.91	0.90
الغابة العشوائية	0.96	0.96	0.96	0.96
XGBoost	0.96	0.96	0.96	0.96

4.4 - الخلاصة :

تشير النتائج إلى أن إعادة توازن البيانات يلعب دوراً مهماً في تحسين أداء نماذج تعلم الآلة في توقع تسرب العملاء. تقنية SMOTE + ENN أثبتت أنها الأكثر فعالية في تحقيق توازن جيد بين الفئات المختلفة، مما أدى إلى تحسين دقة النماذج وأدائها بشكل عام.

5 - الخاتمة :

يعمل نظامنا بفعالية لتحقيق الهدف الرئيسي لتحليل خوارزميات التعلم الآلي المختلفة اللازمة لتطوير نماذج تصنيف التناوب (Churn) للعملاء وتحديد أسباب التناوب من أجل تقديم استراتيجيات وخطط الاحتفاظ بالعملاء. يركز نظامنا على حل مشكلة لا مفر منها وهي تناوب العملاء في قطاع الاتصالات لعدة أسباب. الجانب الأكثر إرباكاً بشأن دوران العملاء هو أن التناوب مستحيل التحكم فيه. علاوة على ذلك، هناك عدة أسباب لفقدان العملاء، بعضها مرني وبعضها الآخر غير مرني. من ناحية أخرى، يجب أن يدرك المشغلون في صناعة الاتصالات أن فقدان العملاء سيحدث عاجلاً أم آجلاً ويجب أن يكونوا مستعدين للرد. في هذا السياق، يلعب نظامنا دوراً مهماً في حل مشكلة تناوب العملاء التي أصبحت ضرورية لبقاء شركات الاتصالات.

المراجع :

- [1] Lalwani P, Mishra MK, Chadha JS. Customer churn prediction system: a machine learning approach. Computing 2022;104:271–94.
- [2] Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques.
- [3] Srinivasan, R., Rajeswari, D., & Elangovan, G. (2023). Customer churn prediction using machine learning approaches.