الجمهوريّة العربيّة السوريّة

وزارة التعليم العالي

المعهد العالي لإدارة الأعمال

السنة الدراسية الخامسة

قسم إدارة العمليات والمعلومات

استخدام تقنيات التنقيب في المعطيات للتنبؤ بنتيجة الحملات التسويقية المباشرة

(توقع ايداعات العملاء في بنك " Banco Português de Investimento ")

إعداد الطالب

محمد ناظم السبيعي

إشراف

د. كادان الجمعة

د. ياسر رحال

العام الدراس

2020/2019

إهداء

إلى من علمني النجاح والصبر... إلى من علمني العطاء بدون انتظار... أبي.

إلى من علمتني وعانت الصعاب لأصل إلى ما أنا فيه... إلى من كان دعاؤها سر نجاحي وحنانها بلسم جراحي... أمي.

إلى جميع أفراد أسرتي العزيزة والكبيرة كل باسمه أينما وجدوا.

إلى ملاكي في الحياة أينما كان.

إلى أصدقائي رفقاء دربي من داخل الجامعة وخارجها.

إلى المشرفين

د. كادان الجمعة د. ياسر رحال

إلى دكاترتي الكرام الذين أناروا دروبنا بالعلم والمعرفة.

إلى كل من يقتنع بفكرة فيدعو إليها ويعمل على تحقيقها، لا يبغي بها إلا وجه الله ومنفعة الناس.

إليكم أهدي ثمرة هذا العمل المتواضع

أرى لزاما على تسجيل الشكر وإعلامه ونسبة الفضل لأصحابه

ما نعلمه قطرة وما لا نعلمه محيط

أولا لله عز وجل على أن هداني لسلوك طريق البحث والتشبه بأهل العلم وإن فالشكر كان بيني وبينهم مفاوز.

كما أخص بالشكر دكاترتي الكرماء

د. وائل خنسة د. طلال عبود د. راضي خازم د. نظرة رحمة د.الاء بركة

والمشرفين الفاضلين على هذا البحث

د. كادان الجمعة د. ياسر رحال

فقد كان حريصا على قراءة كل ما أكتب ثم يوجهني إلى ما يرى بأرق عبارة وألطف إشارة، فله مني وافر الثناء وخالص الدعاء.

كما أشكر السادة الأساتذة وكل الزملاء وكل من قدم لي فائدة أو أعانني بمرجع، أسأل الله أن يجزيهم عني خيرا وأن يجعل عملهم في ميزان حسناتهم.

جدول المحتويات

1-1- مقدمة 9

1-2- مشكلة البحث: 11

1-3- أهداف البحث: 11

1-4- منهجية البحث: 12

1-4- حدود البحث 12

1-5- أهمية البحث 13

1-6- معوقات البحث 14

2-1- الدراسات السابقة 17

2-1-1- الدراسة الثانية (Paulo Cortez) 17

2-1-2- الدراسة السادسة: (جلال الضار) 19

2-1-3- الدراسة الثالثة (Chris Rygielski) (Jyun-Cheng Wang) 21

2-1-4- الدراسة الرابعة (سمية فجخي) 23

2-1-6- الدراسة ال خامسة (راضية نصر الله) (سارة بن ذيب) 25

3-1- التنقيب في البيانات (Data mining) 29

3-1-1- مفهوم التنقيب في البيانات 29

3-1-2- أنواع التنقيب في البيانات 29

3-1-3- أهمية التنقيب في البيانات 30

3-1-4- أهداف التنقيب في البيانات 30

3-1-6- تنقيب البيانات الخاضع للإشراف والتنقيب غير الخاضع للرقابة 32

3-1-7- تقنيات التنقيب في البيانات 33

3-1-7-1- التصنيف classification 33

4-1- التسويق المصرفي 50

4-1-1- مفهوم التسويق المصرفي 50

4-1-2- لأهداف التسويق المصرفي 51

4-1-3- وظائف إدارة التسويق المصرفي 52

4-1-4- أهمية التسويق المصرفي 53

4-1-5- خصائص التسويق المصرفي 53

4-2- التسويق المباشر 55

4-2-1- مقدمة 55
4-2-2- مفهوم التسويق المباشر 55
4-2-3- مميزات وعيوب التسويق المباشر 55
4-2-4- أنواع التسويق المباشر 57
4-3- الودائع البنكية 58
4-3-1- مفهوم الوديعة 58
4-3-2- أنواع الودائع 58
4-4- حاجة البنوك للتنقيب في البيانات 59
5-1- المبحث الأول 61
5-1-1- لمحة عن البنك (Banco Português de Investimento) 61
5-1-1-2- هوية البنك 62
5-1-3- العلامة التجارية 62
5-1-2-1- أداة (weka) 63
5-1-2- شرح محتوى البيانات المدروسة 65
- السمات الخاصة بالعميل 66
- السمات الخاصة بآخر جهة اتصال للحملة الحالية 68
- السمات الخاصة بالحملة التسويقية الحالية والحملات السابقة 69
- السمات الاجتماعية الاقتصادية 70
- متغير الإخراج 71

6-1- المبحث الثاني 72
6-1-1- تصوير البيانات (Data Visualize) 72
- ❖ **المتغيرات الرقمية** 72
- ❖ المتغيرات الفئوية: (categorical) 78

6-1-2- اختيار التقنيات المستخدمة 82
6-1-3- بناء المصنفات 83
6-1-3-1- تحضير البيانات 83
6-1-4- مشاكل التصنيف غير المتوازن 97
6-1-5- التصنيف: (Classification) 99
6-1-5-1- بناء المصنف: 99

2-5-1-6- تقييم أداء المصنف: 99

3-5-1-6- خوارزميات التصنيف: (Classification algorithms) 101

3-5-1-6- اختيار المصنف 107

1-7- النتائج: 109

- **النتائج النظرية للدراسة:** 109
- **النتائج التطبيقية للدراسة:** 110

2-7- التوصيات 111

- **توصية خاصة للبنوك** 111
- **دراسات خاصة للاكاديميين والدراسات المستقبلية** 112

1) مراجع 113

ملخص البحث

يهدف البحث من خلال اعتماده النهج التنبؤي للتنبؤ بالحملات التسويقية المباشرة (الهاتفية) لتوقع إيداعات العملاء في بنك (Banco Português de Investimento) البرتغالي/ بورتو، وكانت فترة جمع البيانات ممتدة من (2008) إلى (2013) وتحديدا من الشهر (4) إلى الشهر (12) في كل سنة، وقمنا بتحليل مجموعة من البيانات تتكون من (23) سمة يتعلق أغلبها بعميل البنك وبعض منها يخص جهة اتصال والحملة التسويقية، وتمت إضافة سمات اجتماعية واقتصادية جديدة.

تم شرح محتوى البيانات وتحليلها من خلال تصوير البيانات (DV) لفهم المعلومات التي تحملها سمات هذه البيانات من أجل البحث في تقنيات التنقيب عن البيانات لاختيار تقنية ملائمة لطبيعة البيانات والمشكلة المدروسة، وقد تم اختيار تقنية التصنيف ردا لطبيعة البيانات التي تحمل سمة (attributes) إخراج مكونة من خيار ثنائي (Yes/No)، وتم اختيار (5) خوارزميات للتصنيف سيتم بناء المصنفات من خلالاهم وهم " الغابات العشوائية (Random Forest)، شجرة القرار (Decision Tree J48)، الانحدار اللوجيستي (logistic regression)، الجار الأقرب (KNN)، وبايز الساذج (naive bayes) "، ثم تم تحضير البيانات المدروسة لإدخالها على خوارزميات التصنيف وقد تم بمرحلة تحضير البيانات المتألفة من عدة مراحل بالاستغناء عن عدة سمات (attributes) من خلال تحليل الارتباط الذي يبحث عن الارتباطات المخفية بين هذه السمات، وقد تم حل مشكلة عدم توازن البيانات باستخدام مبدئ (Random Oversampling)، وبعد عملية التحضير سيتم تطبيق الخوارزميات السابقة عليها لبناء المصنفات واختيار المصنف الأمثل من خلال المقارنة في ما بينهم بعدة معاير وهم (الدقة (accuracy) والإحكام (precision) والحساسية (recall)، وبعدما تمت المقارنة المقارنة بين نتائج المصنفات ال خمس المستخدمة أتضح أن مصنف الغابات العشوائية (Random Forrest) قد حصل تميز بجميع نتائج معايره على المعاير الأخرى من حيث الدقة بمعدل (95.3%)، وإضافة نتائج خاصة بسمات معينة للحصول على نتائج عالية للحملات التسويقية جنبا بجنب مع مصنف الغابات العشوائية (Random Forest).

الفصل الأول

الإطار التمهيدي

يشرح هذا الفصل المشكلة الرئيسية التي دعت إلى انشاء البحث وما هي أهداف وأهمية البحث المدروس، والمنهجية المستخدمة في نطاق هذا البحث، ومن ثم محددات البحث ومعوقاته.

1-1- مقدمة

تكرث علوم البيان خطواتها من خلال التنقيب في المعطيات لاستخراج المعارف الموجودة في البيانات بتقنيات ونماذج تنبؤية ووصفية عديدة للوصول لمعرفة كانت مخبئة داخل البيانات، ومن هنا اردت بناء بحث يجمع بين ثلاثة اختصاصات قد تم الإلمام بها من خلال مرحلتي الجامعية (إدارة العمليات والمعلومات، والتسويق، والمالية) للخروج بدراسة تدعم ما تعلمته خلال سنين دراستي الجامعية، ومنه فقد تم الوصول لدراسة تتنبئ بنتائج بالحملات التسويقية المباشرة لإيداعات العملاء.

تشكل حملات البيع التسويقي استراتيجية نموذجية لتعزيز الأعمال المصرفية، وتستخدم الشركات التسويق المباشر عند استهداف شرائح العملاء من خلال الاتصال بهم لتحقيق هدف محدد أي أن مركزية العميل عن بعد في التعاملات في مركز الاتصال يسهل الإدارة التشغيلية للحملات.

تسمح هذه المراكز بالتواصل مع العملاء من خلال قنوات مختلفة، الهاتف (خط ثابت أو الجوال) كونها واحدة من الأكثر استخداما على نطاق واسع بسوق التشغيل من خلال مركز اتصال يسمى التسويق عبر الهاتف بسبب خاصية البعد يمكن تقسيم جهات الاتصال إلى الداخل والصادر، اعتمادًا على الجانب الذي قام بتشغيل جهة الاتصال (العميل أو جهة الاتصال مركز)، مع كل حالة تطرح تحديات مختلفة (على سبيل المثال، المكالمات الصادرة غالبا ما تعتبر أكثر تطفل). تتيح التكنولوجيا إعادة التفكير في التسويق من خلال التركيز على زيادة القيمة الدائمة للعميل من خلال تقييم المعلومات المتاحة ومقاييس العملاء، وبالتالي السماح لبناء أطول وعلاقات أكثر إحكامًا تتماشى مع طلب الأعمال. أيضا، يجب أن يكون وشدد على أن مهمة اختيار أفضل مجموعة من العملاء، أي أكثر من المحتمل أن يشترك في منتج.

تستخدم أنظمة دعم القرار(DSS) تكنولوجيا المعلومات لدعم الإنسان صنع القرار (agerial) هناك العديد من المجالات الفرعية لـ (DSS)، مثل المجالات الشخصية و (DSS) ذكي ترتبط (DSS) الشخصية بأنظمة صغيرة الحجم لدعم مهمة اتخاذ القرار لمدير واحد، في حين أن (DSS) الذكي يستخدم الفن بتقنيات الذكاء لدعم القرارات.

مفهوم (DSS) آخر ذو صلة هو ذكاء الأعمال (BI)، وهو مصطلح شامل يتضمن المعلومات تقنيات الدعم، مثل مستودعات البيانات والتنقيب عن البيانات (DM)، لدعم صنع القرار

باستخدام بيانات العمل يمكن لـ (DM) لعب دور رئيسي في الشخصية و(DSS) الذكي، مما يسمح بالاستخراج شبه التلقائي للشرح والمعرفة التنبؤية من البيانات الأولية، وعلى وجه الخصوص التصنيف هو أكثر مهام (DM) شيوعًا والهدف هو بناء نموذج قائم على البيانات يتعلم وظيفة أساسية غير معروفة تحدد العديد من متغيرات الإدخال، التي تميز عنصرًا (على سبيل المثال، (عميل البنك)، مع هدف إخراج واحد مصنّف على سبيل المثال، نوع بيع الإيداع المصرفي: ("فشل" أو "نجاح").

في هذه الورقة، نقترح (DSS) شخصي وذكي يمكنه التشغيل الآلي توقع نتيجة مكالمة هاتفية لبيع الودائع طويلة الأجل باستخدام نهج (DM) مثل (DSS) قيمة لمساعدة المديرين في تحديد الأولويات وتحديدها محاضرة العملاء التاليين ليتم الاتصال بهم خلال حملات التسويق المصرفي على سبيل المثال، باستخدام تحليل التأثير الذي يحلل احتمالية النجاح ويترك للمديرين فقط القرار بشأن عدد العملاء للاتصال، ونتيجة لذلك، سيتم تخفيض الوقت والتكاليف لهذه الحملات، أيضا من خلال إجراء عدد أقل من المكالمات الهاتفية الفعالة، وضغوط العميل وسوف يتناقص التطفل المساهمات الرئيسية لهذا العمل.

التركيز على هندسة الميزات، التي تعد جانبًا رئيسيًا في(DM)، ونقترح مؤشرات اجتماعية واقتصادية عامة بالإضافة إلى المزيد من المعلومات سمات العميل المنتج والمنتجات البنكية أحادية الاستخدام، وتتكون من 23 وسمة. في مرحلة النمذجة، وقد تم تقليص المجموعة الأصلية إلى19 ميزة ذات صلة تستخدمها نماذج (DM) نقوم بتحليل مجموعة بيانات تابعة للبنك البرتغالي (Banco Português de Investimento) وتتكون من (41188) سجل، وقد تم جمع البيانات من عام (2008) إلى (2013)، بما في ذلك الآثار الإلكترونية للأزمة المالية العالمية التي بلغت ذروتها في عام 2008. قارنا أربعة نماذج للتنقيب في البيانات (RF/KNN/DT/ Bayes/ LR) باستخدام النماذج التنبؤية للتصنيف، ومن ثم اختيار المصنف الأمثل الذي يعطي أعلى دقة من خلال عدة معايير وهم (الدقة (accuracy) والإحكام (precision) والحساسية (recall).

2-1- مشكلة البحث:

يركز هذا البحث بشكل أساسي على الاستهداف من خلال المكالمات الهاتفية عبر الهاتف لبيع ودائع استثمارية طويلة لأمد خلال الحملة محددة، يقوم عملاء البنك بإجراء مكالمات هاتفية إلى قائمة العملاء لبيع الودائع (الصادرة) أو، إذا كان العميل في الوقت نفسه يتصل مركز الاتصال لأي سبب آخر، يطلب منه الإيداع (داخلي) وبالتالي، فإن النتيجة هي اتصال ثنائي ("غير ناجح أو ناجح")

فإن سؤال البحث التالي ذو صلة، وبل تالي يمكن تلخيص مشكلة البحث من خلال السؤال التالي:

هل يمكن التنبؤ بنتائج الحملات التسويقية (الهاتفية) لتوقع فيما إذا كان العميل سيشترك في وديعة استثمارية طويلة الأجل (نعم أم لا)

3-1- أهداف البحث:

يهدف البحث إلى التنبؤ فيما إذا كان العميل سيشترك في وديعة استثمارية لأجل من خلال الحملات التسويقية المباشرة (الهاتفية) وذلك من خلال الإجابة عن تساؤلات البحث المذكورة في مشكلة البحث، ويمكن تلخيص الاهداف وفق الآتي:

1) إدارة الحملات التسويقية بطريقة فعالة.
2) تخفيض تكلفة الحملات التسويقية من خلال استهداف الشرائح الفعالة.
3) زيادة الإيرادات البنك من خلال الحملات التسويقية.
4) اكتساب عملاء جدد وضمهم للشرائح المستهدفة.
5) تصنيف العملاء الموجودين حاليا للتنبؤ بحاجات العملاء ذاتهم والعملاء المرتقبين.
6) زيادة ايداعات العملاء من خلال الحملات التسويق الهاتفية.
7) التأكد من رضا العميل من خلال تلبية احتياجاته.

4-1- منهجية البحث:

المنهجية المستخدمة هي المنهجية التنبؤية والتي يعتمد عليها علم التنقيب البيانات (Mining Data) في تنبوءاته ما يعرف بالنموذج التنبؤي (Model Predictive) وهو يرتكز على نموذج إحصائي يعتمد على عوامل متغيرة وهي غالبا ما تؤثر على السلوك المستقبلي والنتائج ويتم بناء هذا النموذج بعدة طرق (methods) والتي تقوم على نظريات الاحصاء والمعادلات الرياضية ومن اهم هذه الطرق على سبيل المثال لا الحصر: التصنيف (Classification)، الانحدار (Regression)، القرارات شجرة (Decision Tree)، التجميع(Clustering).

4-1- حدود البحث

تتلخص محددات البحث من خلال الحدود الزمانية والمكانية وفق الآتي:

- حدود زمانية: تم إعداد البحث خلال المدة الزمنية الممتدة ما بين 2020/6/15 و 2020/8/1

- حدود مكانية: تمثلت الحدود المكانية لهذه الدراسة في بنك

(**Banco Português de Investimento**) / بورتو

5-1- أهمية البحث

يمكن توضيح أهمية البحث من خلال شقين أساسيين، الأهمية النظرية من البحث والأهمية التطبيقية منه وفق الآتي:

- أهمية النظرية:

تتجلى الأهمية النظرية للبحث من خلال توضيح المراحل لتي مرت بها عملية بناء النموذج التنبؤي للبنك المدروس بشفافية وتعتبر هذه الدراسة ذو أهمية كبيرة بالنسبة لمجال علم البيانات والتنقيب ب البيانات، فسوف يتم توضيح العديد من التعاريف والمصطلحات المتعلقة بهذا البحث، فيمكن أن يكون هذا البحث مرجعا للدراسات القادمة في هذا المجال لما يثيره من قضايا وتساؤلات يمكن أن تؤخذ بعين الاعتبار حيث سيتم تقديم توصيات تساهم في تحسين النماذج التنبؤية لحملات التسويق المصرفية التي تعتمد على ذكاء الأعمال والتنقيب في البيانات الأخرى الشبيهة والمماثلة.

- أهمية تطبيقية:

تتجلى الأهمية التطبيقية للبحث من خلال النموذج في التنبؤ بلعملاء الذين سيشتركون في وديعة استثمارية ام لا "وهذا ما سيتم بنائه في الإطار العملي" بحيث يساهم وبشكل كبير في تحسين الحملات التسويقية المستقبلية من خلال مساعدة المديرين في تحديد الأولويات وتحديدها محاضرة العملاء التاليين ليتم الاتصال بهم خلال حملات التسويق المصرفية المستقبلية، ومنه يتم استبعاد المكالمات التي تنبئ بها المصنف بأنها غير ناجحة أي (ليست ستشترك بوديعة) ومنه تسليط ضوء الحملات التسويقية القادمة للعملاء الأكثر احتمالا من خلال بياناتهم بأنهم سيشتركون بوديعة ومن ثم نقل كفاءة الحملات التسويقية من مكان الى أخر، حيث سيتم تقديم نتائج وتوصيات تساهم في تحسين أنظمة التنبؤ للشركات والبنوك التي تعتمد على الحمالات التسويقية المباشرة على (الهاتف) بشكل عام وفي هذا المجال.

6-1- معوقات البحث

1) صعوبة الحصول على دراسات سابقة مشابهة لمنهجية البحث المستخدمة، كون المشاريع المشابهة عادة ما تتسم بالحماية الفكرية وبل تالي المعلومات التي تحويها تبقى محدودة.
2) بعد الشركات والبنوك عن علم البيانات ونتائجه وتأثيره.
3) عدم وجود مراجع تستخدم نفس الأدوات المستخدمة.
4) ضيق الوقت اللازم لإنهاء المشروع.
5) الوضع الراهن بسبب جائحة كورونا وأثارها.

سوف نتحدث بالفصول القادمة عن:

الفصل الثاني:

سنتحدث عن الدراسات السابقة التي ساعدت في بناء البحث.

الفصل الثالث:

سنتحدث عن ماهية التنقيب في المعطيات، وأهميته، وأهدافه، وخطوات استخراج المعرفة من البيانات، وأهم التقنيات اللازمة لبناء البحث.

الفصل الرابع:

سنتحدث بشكل مفصل عن التسويق البنكي، وعن التسويق المباشر، والودائع البنكية، وحاجة البنوك للتنقيب في البيانات.

الفصل الخامس:

ينقسم الفصل التالي إلى مبحثين:

المبحث الأول: سيتحدث بلمحة سريعة عن البنك الذي أجريت عليه دراسة هذا البحث، ثم سنتحدث عن الأدوات المستخدمة بمجال التنقيب عن البيانات والتي سوف نحتاج اليها في هذا البحث، ومن ثم سيتم تقديم محتوى البيانات التابعة للبنك وتفسيرها وشرحها.

أما المبحث الثاني: فسيتحدث عن تصوير البيانات (Data Visualize)، وتجهيز البيانات وحل مشكلة عدم التوازن من أجل استخدام تقنيات التنقيب في التصنيف (Classification) ومن ثم اختيار المصنف الأفضل

الفصل السادس:

النتائج، والتوصيات، والمراجع.

الفصل الثاني

الدراسات السابقة

في هذا الفصل سوف نتناول أبرز الدراسات التي ساهمت في بناء هذا البحث وما قدمته هذه الدراسات من استفادة ومعلومات نظرية وعملية قيمة

1-2- الدراسات السابقة

1-1-2- الدراسة الثانية (Paulo Cortez)

(نهج التنقيب عن البيانات للتسويق الدولي باستخدام حزمة rminer وأداة R)

- **الملخص:**

بسبب الأزمة المالية العالمية، أصبح الائتمان في الأسواق الدولية أكثر تقييدًا، وتحويل الانتباه إلى العملاء الخارجيين قاد هذا السياق إلى طلب معرفة سلوك الزبون تجاه البضائع الخارجية وخاصةً الاستجابة لحملات التسويق الدولية، ويصف هذا العمل نهجًا لاستخراج البيانات لاستخراج المعرفة القيمة من الأحدث بيانات حملة التسويق الدولية. وقد استرشد هذا النهج من خلال –CRISP– تم تنفيذ منهجية إدارة البيانات وتحليل البيانات باستخدام حزمة (rminer) وأداة (R)، وتم اختبار ثلاثة نماذج تصنيف (مثل أشجار القرار ومصنف بايز ودعم آلات المتجهات) ومقارنتها باستخدام معيارين ذوي صلة: (ROC) ومنحنى الرفع تحليل. بشكل عام، حصلت آلة المتجهات الداعمة على أفضل النتائج وحساسة تم تطبيق التحليل لاستخراج المعرفة المفيدة من هذا النموذج، مثل أفضل الشهور للترويج وتأثير ثقافة البلدان للحملة على المخرجات.

- **النتائج:**

في التسويق الدولي، يمكن أن تكون نتائج التقارير من حملات التسويق المنفذة تستخدم لتحديد اتجاهات سلوك العميل الخارجي ورغم صعوبة قياسه ألا أن المديرون يتحولون من التقليدية تحليل إحصائي نحو تقنيات (DM) أكثر تعقيدا، من أجل استخراج مفيدة المعرفة من البيانات الخام، تُظهر حالة الطلب الواقعية المعروضة هنا أنه من الممكن تحليل بيانات سلوك الزبون من خلال التسوق عبر المواقع الالكترونية باستخدام مواقع مفتوحة المصدر، على وجه الخصوص، باستخدام بيئة (R) وحزمة (rminer) استرشد مشروع (DM) هذا بمنهجية (CRISP–DM)، تحت ثلاث سنوات التكرارات. تم استكشاف ثلاث تقنيات: DM (NB وDT وsvm) وعموما، لكلا (ROC) وتحليل منحنى الرفع، حصل (SVM) على أفضل النتائج التنبؤية.

- **الاستفادة من الدراسة:**

تمت الاستفادة من هذه الدراسة كونها تتبع النهج التنبؤي، وتم الإمعان في محاور الدراسة النظرية التي تتكلم على علم التنقيب في البيانات وتقنياتها زيادة أن صاحب الدراسة قد تبحر في التنقيب

عن البيانات مما راق لي لسهولة ترتيب المعلومات عند قرأتها للمتلقي، أما ما يوجد في التقنيات التي استخدمها مثل (أشجار القرار ومصنف بايز الساذج) فقد رسخت المعلومات التي اكتسبتها من الدراسات الأخرى عن شجرة القرار وزودت معرفتي بكيفية فهم آلية عمل مصنف بايز ومصنف الآلات الداعمة ومتى تستخدم هذه المصنفات، وكيفية تطبيع وتعميم البيانات خطوة بخطوة، مجملا الدراسة رائعة بالنسبة للتقنيات وخطوات البيانات وعلم التنقيب عن المعطيات، ولم استفد من القسم التسويقي لأنه بعيد عن مجال دراستي هذه.

2-1-2- الدراسة السادسة: (جلال الضار)

(تصميم نموذج نظام دعم القرار لإدارة الموارد البشرية بالاعتماد عمى تقنيات الذكاء الصنعي)

- **الملخص:**

يرمي البحث بشكل أساسي نحو تحقيق الأهداف الثلاثة التالية:

1) دراسة وتحليل كل من فعاليتي تقيم أداء العامليين والتفقد اليومي في الجمهورية السورية كنماذج لفعاليات الكوادر البشرية للاستفادة منه بياناتنا ضمن نطاق الذكاء الصنعي.
2) انتهجت الدارسة أسلوب البحث التحليلي التطبيقي من خلال دارسة وتحميل بعض فعاليات الكوادر البشرية (عملية تقييم أداء العاملين، عملية التفقد اليومي) وأدوات التنقيب في المعطيات (التصنيف، اختيار، المعايير، العنقدة) ومن ثم اقتراح النموذج الملائم لطبيعة العمل في الجمهورية العربية السورية بالاستناد إلى معطيات حقيقية تم جمعيا من بعض المؤسسات والشركات ومناقشة النتائج التي تم الحصول عليها.
3) اقتراح نموذج تطبيقي للإدارة الذكية لموارد البشرية يستند إلى التنقيب في المعطيات لمعالجة عينة البيانات التي تم جمعيا بغية استكشاف المعرفة الكامنة داخل هذه المعطيات بما يدعم عملية اتخاذ القرارات المتعمقة بالكوادر البشرية في مؤسسات الجمهورية العربية السورية ومن ثم مقارنة أداء التقنيات التي تم استخدامها وبيان مدى ملاءمتها ليذا النوع من المعطيات.

- **النتائج:**

1) تقدم أدوات التنقيب في المعطيات إمكانات واسعة لتطبيقات الذكاء الصنعي لشتى أنواع البيانات ومن بينيا معطيات فعاليات الكوادر البشرية مما يسمح بتحميل متقدم لهذه البيانات يؤسس أطر وآليات الإدارة الذكية للموارد البشرية.
2) (Weka) يشكل تطبيق أداة فعالة للاستفادة من تقنيات التنقيب في المعطيات من خلال ما يمنحه هذا التطبيق مكانات متنوعة، حيث ويسمح هذا التطبيق بتجهيز البيانات وفق صيغ متعددة قابلة للتعديل والحفظ والنقل بسهولة ومن ثم استخدامها وفق شتى تقنيات التنقيب في المعطيات من التصنيف والعنقدة وقواعد الترافق وانتهاء بل معاير بالإضافة لما يوفره هذا التطبيق من الإمكانات المرئية (Visualization) الأمر الذي يسمح بتسهيل العمل ومعاينة النتائج ومقارنتها بشكل أكثر فعالية.

3) باستخدام أداة التصنيف يمكن استخدام نتائج عملية تقييم أداء العاملين في بناء نماذج قادرة على التنبؤ بالوظيفة المناسبة لموظف بناء على تقييمه الحالي وذلك مع الفرض بأن التقييم ذو مصداقية وشفافية وبعيد عن العلاقات والأهواء الشخصية.
4)) إن أداء كل من خوارزميات التصنيف (MLP, SVM, J48, Naïve Bayes) كان جيدا في التعامل معً هذا النوع من المعطيات ل بينما كان أداء خوارزمية (KNN) ضعيفا مما يستدعيً توسع أكثر في دارسة مدى إمكانية الاستفادة من هذا المصنف لمتعامل مع هذا النوع من البيانات.
5) يشكل التنقيب في المعطيات مرتكزا أساسيا في الإدارة الذكية للموارد البشرية.
6) تشكل معطيات فعاليات الكوادر البشرية منبعا مما يستوجب إعطاءها الأهمية المطلوبة

- **الاستفادة من الدراسة:**

من بين جميع الدراسات هذه الدراسة عي الأغنى من حيث المعلومات فقد كان محتوى الدراسة يتمحور حول دعم القرارات من خلال التنقيب عن البيانات فقد قدم تسلسل رائع لعملية التنقيب عن البيانات من الخطوة الأولى حتى شرح الأدوات التي تعمل التقنيات من خلالها، مع شرح مصنف (J48) التابع لأشجار القرار والذي سوف يستخدم بدراستي الحالية، وإعادة النظر من خلال سرده للمصنفات التالية(أشجار القرار/ مصنف بايز) على رغم معرفتي بهم من مراجع أخرى لكن الدراسة غنية جدا بتفاصيل لا يقدمها إلا دكتور في جامعة، واستفدت من تعلم آلية عمل مصنف الجار الأقرب (KNN) الذي ذكرها كتقنية ومن ثم طبق عليها في البيانات المتاحة لديه، وخوارزمية (K–means) التابعة لخوارزميات العنقدة، وقد استفدت أيضا من شرح أداة (Weka) التي قد شرحت بل كامل، وعموما فقد اعتمدت فهم شرح الأداة من دراسته مقارنة بموقع الأداة الرسمي وقد ذكر واجهة الأداة وكيفية استخدامه مع التحدث عن مصفوفة الشك الخاصة بالأداة، مجملا هذه هي من أفضل الدراسات التي قد زودتني بكم من المعلومات لا يوصف من شرح طريقة عمل كل تقنية لكن ليس تقنية واحدة (مثل: التصنيف) بل أكثر من عدة تقنية، زيادة على ذلك خطوات استخراج المعرفة من البيانات وشرح مفهوم الانحدار، وشرح ارتباط علم البيانات بدعم القرار الذي تفتقره باقي الدراسات.

3-1-2- الدراسة الثالثة (Chris Rygielski) (Jyun-Cheng Wang)

(تقنيات استخراج البيانات لإدارة علاقات العملاء)

- **الملخص:**

جعلت التطورات التكنولوجية من تسويق العلاقات حقيقة واقعة في السنوات الأخيرة. وقد جعلت تكنولوجيات مثل تخزين البيانات، واستخراج البيانات، وبرمجيات إدارة الحملات، إدارة علاقات العملاء مجالًا جديدًا حيث يمكن أن تكتسب فيها الشركات ميزة تنافسية. بشكل خاص من خلال استخراج البيانات واستخراج المعلومات التنبؤية الخفية من قواعد البيانات الكبيرة ويمكن للمؤسسات تحديد العملاء القيّمين، والتنبؤ بالسلوكيات المستقبلية، وتمكينها من اتخاذ قرارات استباقية وقائمة على المعرفة. أصبحت التحليلات المؤتمتة والموجهة نحو المستقبل ممكنة عن طريق استخراج البيانات تتجاوز تحليلات الأحداث الماضية التي توفرها عادةً الأدوات الموجهة للتاريخ مثل أنظمة دعم القرار. تجيب أدوات التنقيب عن البيانات على أسئلة الأعمال التي كانت في الماضي تستغرق وقتًا طويلاً بحيث لا يمكن متابعتها. ومع ذلك، فإن الإجابات على هذه الأسئلة تجعل إدارة علاقات العملاء ممكنة. توجد تقنيات مختلفة بين برامج استخراج البيانات، لكل منها مزاياه وتحدياته لأنواع مختلفة من التطبيقات.

- **النتائج:**

يقوم استخراج البيانات بإجراء تحليل يكون معقدًا للغاية ويستغرق وقتًا طويلًا للإحصائيين، ويصل إلى شذرات غير معروفة من قبل من المعلومات يتم استخدامها لتحسين الاحتفاظ بالعملاء، ومعدلات الاستجابة، والجذب، والبيع المتقاطع من خلال التنفيذ الكامل لبرنامج CRM، الذي يجب أن يشمل استخراج البيانات، تعزز المؤسسات الولاء المحسن، وتزيد من قيمة عملائها، وتجذب العملاء المناسبين مع تفاعل العملاء والشركات بشكل أكثر تواتراً، سيتعين على الشركات الاستفادة من CRM والتقنيات ذات الصلة لالتقاط وتحليل كميات هائلة من معلومات العملاء. تتمتع الشركات التي تستخدم بيانات العملاء وموارد المعلومات الشخصية بفاعلية في تحقيق النجاح. ومع ذلك، إن الشركات والمديرين هم الذين يحددون كيفية استغلال البيانات التي تم جمعها، ومن واجب الشركات أيضًا تنفيذ سياسة الخصوصية الخاصة بها من أجل إنشاء علاقات جيدة مع العملاء والحفاظ عليها لبناء الثقة والحفاظ عليها، إن طبيعة الثقة هشة لدرجة أنها بمجرد انتهاكها تختفي. تركز حلول CRM الحالية بشكل أساسي على تحليل

معلومات المستهلك من أجل الفوائد الاقتصادية، والقليل جدًا من اللمسات لضمان الخصوصية. نظرًا لأن مشكلات الخصوصية تصبح مصدر قلق كبير للمستهلكين، فمن المؤكد أن الحل المتكامل الذي يبسط ويعزز العملية الكاملة لإدارة علاقات العملاء سيصبح أكثر ضرورة.

- **الاستفادة من الدراسة:**

تمت الاستفادة من هذه الدراسة في جانب النظري فقط حيث تم الإلمام إلى المفاهيم الرئيسية للتقنيات والمفاهيم الكثيرة المسرودة عن علم التنقيب عن المعطيات وحاجة جميع القطاعات اليه، مع تفصيل الخطوات التي يجب أن تطرق على البيانات لاستخراج المعرفة منها فقد شرحت بطريقة سهلة ومفيدة إلى جانب باقي الدراسات.

4-1-2- الدراسة الرابعة (سمية فجخي)

(أثر التسويق المصرفي في تحقيق الميزة التنافسية للبنوك)

- **الملخص:**

تهدف هذه الدراسة إلى إبراز أثر التسويق المصرفي في تحقيق الميزة التنافسية للبنوك التجارية خاصة لأنواع وخصائص التسويق المصرفي، وعناصر المزيج التسويقي التي تمثل أهم العوامل المؤثرة في تحقيقها، وبغرض توضيح ذلك تم إسقاط الدراسة على بنك القرض الشعبي الجزائري (وكالة عين مليلة) الاختبار أثر التسويق المصرفي في تحقيق الميزة التنافسية.

- **النتائج:**

1) التسويق المصرفي يساهم في زيادة أرباح البنك.
2) تعدد وتنوع استراتيجيات تطوير الميزة التنافسية في البنوك.
3) أن إدارة التسويق كهيئة قائمة بذاتها لم نلمس تواجدها بالوكالة.
4) هناك تطبيق جزئي لعناصر المزيج التسويقي بالوكالة.
5) الخدمات التي يقدما البنك ال تختلف كثيرا على ما تقدمه البنوك الأخرى باستثناء تميزه في مجال الخدمات الالكترونية.
6) نظرا لغياب المنافسة في القطاع العام الجزائري، وكونها تحدد مسبقا من قبل الوصاية، تم تسجيل غياب عامل السعر كعنصر جذاب وداعم لتحقيق الميزة التنافسية.
7) نقص الاستغلال الأمثال للأجهزة الآلية داخل الوكالة وغياب شبكات الرابط عند معالجة
8) العمليات البنكية مآبين البنوك؛ مما يعطل سير مصالح الزبون وغياب الدقة في أداء. العمليات وهذا ما يفسر التأثير الضعيف لبعد العمليات البنكية في دعم الميزة التنافسية للبنك.
9) فيما يخص الترويج لمسنا اهتمام ضعيف وغير كافي، وغياب أي برنامج ترويجي مميز للوكالة وهذا ما أثبتته الدراسة الإحصائية وهو وجود أثر ضعيف للترويج في دعم الميزة التنافسية لهذا البنك.
10) ما فيما يخص التوزيع لاحظنا أن مكان اختيار الفرع مناسب وملائم ومحدد بدقة، وهذا ما لاحظناه من خلال الدراسة الإحصائية التي تقر بوجود تأثير قوي لبعد التوزيع على القدرة التنافسية للبنك.

- **الاستفادة من الدراسة:**

قد تم الاستفادة من هذه الدراسة عبر كم المعلومات الهائل الموجود بها الذي يتكلم عن شرح مفهوم التسويق المصرفي وذكر كيفية نشأته وحاجت جميع المصارف إلى التسويق، ومن ثم انتقلت إلى أنواع التسويق المصرفي وتمركزت الفائدة المأخوذة منها عن فهم التسويق المصرفي عبر (الهاتف) الذي قد طرح في مشكلة البحث التي ادرسها، ومن ثم تطرقت طالبة البحث إلى خصائص التسويق الهاتفي وأهميته وأهدافه التي من خلالاها يحقق المصرف الميزة التنافسية لديه وكذلك لباقي أنواع التسويق والحملات التسويقية المصرفية، واستفدت من عمليات البنك التي قد فصلت بشكل دقيق وقد ركزت استفادتي على عمليات الإيداع وآلية عملها لدى المصارف وما هي أنواع الودائع وما يهمني من الأنواع هو الودائع الاستثمارية فقد فاضت الطالبة في الحديث عنه إلى جانب باقي الأنواع، وقد استفدت من بحثها من خلال لأثر الفعلي لقيمة التسويق المصرفي بشتى أنواعه ونشاطاته على المصارف وميزتها التنافسية أمام باقي البنوك وقد رسخت تأثيرها من خلال النتائج التي اظهرها البحث، فمجملا إن البحث غني جدا بالمعلومات القيمة التي تدعم بحثي نظريا، لكن اقتصر البحث على التكلم عم التسويق المصرفي بشكل خاص وهمش قليلا موضوع الحملات وهذا ليس ضعفا فالطابة قد بنت بحثها على أساس التسويق المصرفي أي أنواع التسويق المصرفي، وكانت طريقة صياغتها لمفهوم التسويق المصرفي جميلة وقد اعتمدت منها في بحثي هذا مرورا بعمليات المصارف وآلية عملها.

6-1-2- الدراسة ال خامسة (راضية نصر الله) (سارة بن ذيب)

(حملات التسويق المصرفي وتحديات تكنلوجيا الإعلام والاتصال)

- **الملخص:**

يمثل التسويق المصرفي ثقافة راقية ونشاط متخصص في مجال التعاملات المصرفية حيث يضع العملاء في قائمة أولوياته من خلال بناء صيغ خلاقة لتحقيق مستويات أعلى من اشباع حاجاتهم الحالية والمستقبلية، ومن أجل ذلك يهتم مسؤولي الوظيفة بتبني المفاهيم الحديثة للتسويق ومواكبة تطورات تكنلوجيا الاعلام والاتصال والتي أحدثت بدورها تغيرات جذرية على نشاط البنك حيث غير البنك مختلف أدواته مثل تبني الصيرفة الالكترونية والتي تحظي بقبول واسع لدى العملاء، وكنتيجة للدراسة وعلى وجه التحديد بنك الفلاحة والتنمية الريفية، والذي كان بدوره من البنوك الجزائرية التي تبنت التوجهات الحديثة التي أفرزتها تكنلوجيا الاعلام والاتصال من أجل التحسين المستمر للخدمات لتوسيع قاعدة زبائنه، فقد قام البنك بتطوير الجانب التقني والجانب البشري له عن طريق ادخال استخدام هذه التكنولوجيات في تنفيذ العمليات البنكية وغيرها من الخدمات الأخرى التي يوفرها البنك لعملائه، حيث أصبحت وسائل الدفع الحديثة من اهم المعاملات المصرفية لديه، وهذا ما جعله من المصارف الرائدة على المستوى المحلي.

- **النتائج:**

التسويق المصرفي هو مجموعة من الأنشطة المتخصصة والمتكاملة التي توجه من خلال موارد المصرف وإمكانياته، تستهدف تحقيق مستويات أعلى من الإشباع لحاجات ورغبات العملاء الحالية والمستقبلية.

يشتمل المزيج التسويقي المصرفي عناصر أساسية تشكل مجموعة نشاطات وحملات تسويقية مصرفية وتتمثل في: المنتج البنكي، والسعر، والترويج، الأفراد الدليل المادي كذلك عملية تقديم الخدمة، وعمليات التسويق (الهاتفي) و(الالكتروني) وعملية الوضع والدمج الحسن لهذا المزيج يمثل جل اهتمامات التسويق المصرفي.

أما عن الدراسة الميدانية لبنك الفلاحة والتنمية الريفية وكالة تبسة (488) ومن خلال المقابلة التي أجريت، والتحليل للاستبيانات الموزعة على أفراد العينة تم استخلاص ما يلي:

يعتبر بنك الفلاحة والتنمية الريفية من أهم المصارف العمومية التي حاولت جاهدة في تطوير الخدمات المصرفية الالكترونية (E-Banking)، وهذه الخدمة تسمح للعميل بفحص الحساب الكترونيا أي باستخدام الإنترنيت وذلك من خلال 24/24 ساعة و7/7 أيام، وتعتبر هذه الخمة نقطة تحول من مستوى تقديم الخدمات وعلاج الكثير من المشاكل مساهمة البنك في التبني لأدوات الدفع الالكتروني في خلق نطاق أوسع للدفع محليا ودوليا.

وقد تبنت نتائج النموذج الإحصائي والذي اعتمد فيه الاستبيان كأداة، أنه يوجد علاقة إيجابية، لها أثر ذو دلالة إحصائية لتكنلوجيا الاعلام والاتصال وهو المتغير المستقل على الوظيفة التسويقية في بنك الفلاحة كمتغير تابع.

- **الاستفادة من الدراسة:**

قد تمت الاستفادة من هذه الدراسة من خلال ربط الحملات التسويقية في التسويق المصرفي وتطبيقاته الخدمية على العملاء لمعرفة احتياجاتهم ورغباتهم فقد ركزت الدراسة أولا على التسويق المصرفي بشكل عام، ثم تكلمت عن المزيج التسويقي الخاص به ومن ثم ربط الحملات التسويقية وتطبيقاتها في المصرف الجزائري لتضيف عليه الميزات من خلالها، فقد تمت الاستفادة من فهم ماهية الحملات التسويقية المباشرة وغير المباشر والفرق بينهم، وكيفية امتداد الحملة من الالف إلى الياء، من أهداف وخصائص وأهمية إلى جانب حملات التسويق الالكترونية التي بني عليها هذا البحث مشكلته للوصل إلى نتائج باهرة وإعطاء البنك ميزات هامة على المستوى المحلي من خلال تطبيقية لحملات التسويق الالكترونية ودعم الدفع الالكتروني، وكانت نتائج البحث تأكد حاجة المصارف للتسويق المصرفي وخدماته بشتا الأنواع لكنه ركز على الحملات الالكترونية، مجملا الدراسة قد زودتني بما ينقصني عن منظومة التسويق المصرفي التي كانت تنقص البحث السابق وها قد أتممت استفادتي من خلال هاذه الدراسة بل إضافة للدراسة السابقة (أثر التسويق المصرفي في تحقيق الميزة التنافسية للبنوك) في مجال التسويق المصرفي بشكل كامل مع الشرح المفصل، فقد رممت كل ما احتاجه بخصوص الحملات التسويقية المصرفية وما تقدمه البنوك من (إيداعات و....).

الفصل الثلث

الإطار النظري

في هذا الإطار سوف ندرس نظرياً ماهية التنقيب في البيانات وما يحتويهما من تقنيات ومفاهيم ومصطلحات تخدم هذا البحث وسوف يتم التحدث عنها كالتالي:

مفهوم التنقيب في البيانات

أنواع التنقيب في البيانات

أهمية التنقيب في البيانات

أهداف التنقيب في البيانات

مراحل استخراج المعرفة والتنقيب بها

تنقيب البيانات الخاضع للإشراف مقابل التنقيب غير الخاضع للرقابة

تقنيات التنقيب في البيانات

تمهيد

أدى الانتشار الواسع لتقنية المعلومات وسهولة إتاحتها إلى تضخم حجم المعلومات بصورة استباقية لم يشهدها التاريخ من قبل، مما جعل من قضية البيانات الضخمة على الإنترنت مثاراً للجدل، من حيث جدوى وجودها بهذه الصورة العشوائية، وعندما نتحدث عن البيانات الضخمة، فإننا نتحدث عن كميات لا يمكن تخيلها من البيانات متعددة الأنواع والمصادر بحجم يصل إلى المئات من التيرابايت أو حتى البيتابايت (البيتابايت هو الرقم واحد متبوعا بـ 15 صفر)، وهنا نتساءل ما أهمية هذه البيانات في ظل وجود حقيقة تشير إلى أن المعلومات المنظمة من هذه البيانات لا تشكل إلا جزءً ضئيلاً يصل إلى 10% مقارنه بالمعلومات غير المنظمة والتي تشكل الباقي ذلك أدي إلى ازدياد الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، فالأساليب التقليدية والإحصائية لا تستطيع أن تتعامل مع هذا الكم من الهائل لذا تستخدم أدوات ذكية لمعالجة هذه البيانات.

من هنا ظهر ما يسمى باستخراج البيانات (Data Mining) كتقنية تهدف إلى استنتاج المعرفة من كميات هائلة من البيانات، تعتمد على الخوارزميات الرياضية والتي تعتبر أساس التنقيب عن البيانات وهي مستمدة من العديد من العلوم مثل علم الإحصاء والرياضيات والمنطق وعلم التعلم، والذكاء الاصطناعي والنظم الخبيرة، وعلم التعرف على الأنماط، وعلم الآلة، والتنقيب عن البيانات وغيرها من العلوم والتي تعتبر من العلوم الذكية وغير التقليدية.

1-3- التنقيب في البيانات (Data mining)

1-1-3- مفهوم التنقيب في البيانات

هناك الكثير من التعريفات لمفهوم تنقيب البيانات، وقد تم اختيار التعريفات التالية:

- هي النشاط الذي يقوم باستخراج المعلومات المتواجدة في كميات كبيرة من البيانات، بهدف البحث عن أنماط معرفية واكتشاف الحقائق الخفية الواردة في قواعد البيانات.[1]
- عملية البحث داخل كميات كبيرة من البيانات للكشف عن العلاقات التي لم يتم كشفها سابقا بين عناصر البيانات؛ والمعروفة أيضا باسم اكتشاف المعرفة في قواعد البيانات (KDD).[2]
- تحليل البيانات المتواجدة في قواعد البيانات باستخدام الأدوات التي تبحث عن الاتجاهات أو البيانات التي لا معنى لها، واستخراج معلومات ضمنية، لم تكن معروفة سابقا، ويمكن أن تكون مفيدة.
- عملية تحليل البيانات من وجهات نظر مختلفة واكتشاف الاختلالات والأنماط والارتباطات الموجودة في مجموعات البيانات التي تكون ثاقبة ومفيدة للتنبؤ بالنتائج التي تساعدك على اتخاذ قرار صائب.
- عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة)، لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات.[3]
- هي عملية بحث محوسب ويدوي عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة.

2-1-3- أنواع التنقيب في البيانات

- **التنقيب الاستشرافي:** ينتج عنه نموذج عن النظام الذي تصفه البيانات المستخدمة في التنقيب.
- **التنقيب الوصفي:** ينتج عنه معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب[4].

---

[1] (Palace, 1996)
[2] (han, kamber, & pei, 2000)
[3] (الضاهر، 2014)
[4] (dean، 2014)

3-1-3- أهمية التنقيب في البيانات

للتنقيب في البيانات أهمية كبيرة في الكثير من التطبيقات الحقيقة حول العالم وأهمها القدرة على التعامل مع المشاكل المعقدة، والقدرة على اكتشاف المعلومات المثيرة للاهتمام والغير متوقعة، واستخراج العديد من البيانات المخفية، والقدرة على التعلم الذاتي، وإمكانية استخدام التجارب والأخطاء من الماضي لتحسين نوعية النماذج تلقائيا، والتعرف على مسارات البيانات المخفية وهو ما يؤهله لتقديم العون والمساعدة في بناء التنبؤات المستقبلية، واستكشاف السلوك، والاتجاهات مما يسمح لاتخاذ القرارات الصحيحة وفي الوقت المناسب. [5]

3-1-4- أهداف التنقيب في البيانات

- من أجل تعليل بعض الظواهر المرئية.
- من أجل التثبت من نظرية ما.
- من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة[6].

3-1-5- مراحل استخراج المعرفة والتنقيب بها

تظهر عملية استخراج المعرفة في الشكل 1.4 كتسلسل تكراري للمراحل التالية:

1) **تنظيف البيانات** (لإزالة الضوضاء والبيانات غير المتسقة).
2) **دمج البيانات** (حيث يمكن دمج مصادر بيانات متعددة).
3) **اختيار البيانات** (حيث يتم استرداد البيانات ذات الصلة بمهمة التحليل من قاعدة البيانات).
4) **تحويل البيانات** (حيث يتم تحويل البيانات ودمجها في أشكال مناسبة للتعدين من خلال إجراء عمليات الملخص أو التجميع).[7]

---

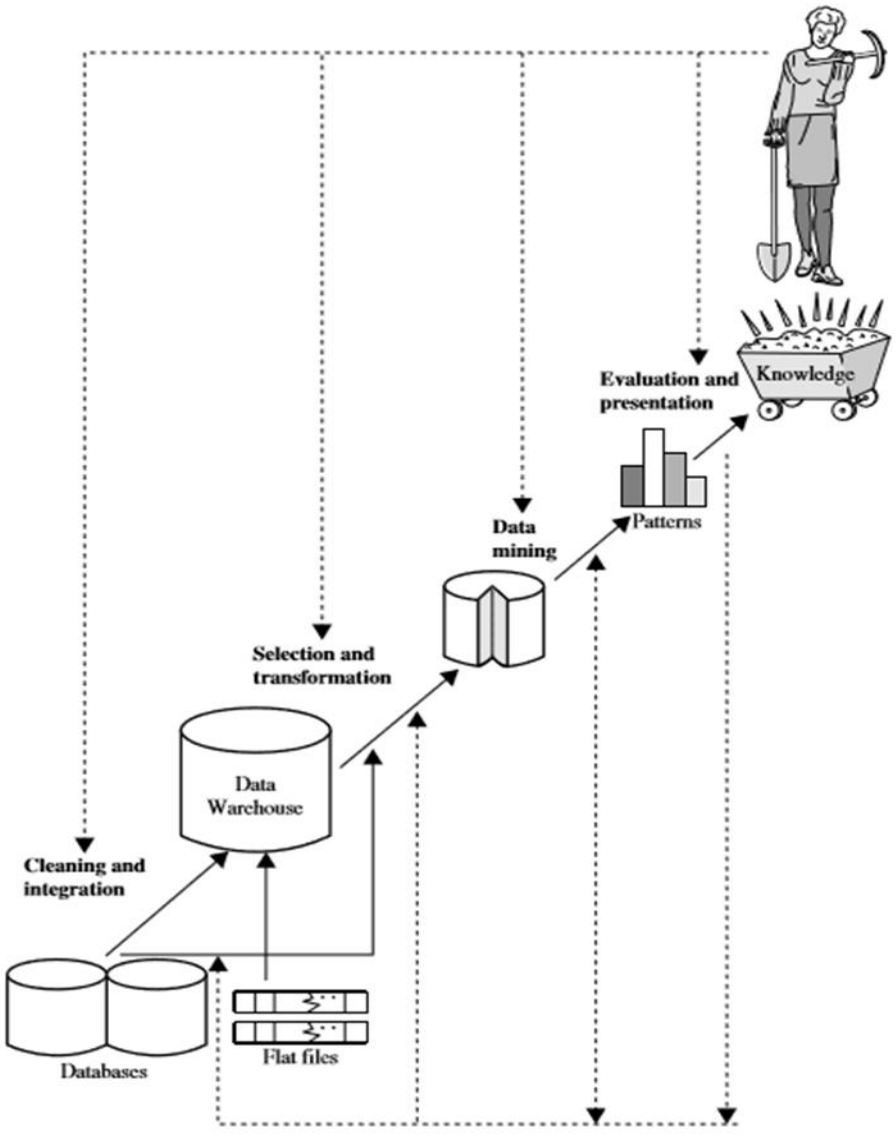
[5] (han, kamber, & pei, 2000)
[6])عبيد, 2019(
[7] (Palace, 1996)

5) **استخراج البيانات** (عملية أساسية حيث يتم تطبيق الأساليب الذكية لاستخراج أنماط ونماذج البيانات).

6) **تقييم النماذج** (لتحديد النماذج التي تمثل المعرفة بناءً على المقاييس المهمة).

7) **عرض المعرفة** (حيث يتم استخدام تقنيات التصور وتمثيل المعرفة لتقديم المعرفة للمستخدمين).[8]

الشكل (1)



Data mining as a step in the process of knowledge discovery.

المراحل من 1 إلى 4 هي أشكال مختلفة من المعالجة المسبقة للبيانات، حيث يتم تحضير البيانات للتنقيب، ومن ثم يتم اختيار التقنية المناسبة للتنقيب فب هذه البيانات.

[8] (han, kamber, & pei, 2000)

6-1-3- تنقيب البيانات الخاضع للإشراف والتنقيب غير الخاضع للرقابة

تتضمن الاستراتيجيتان الأكثر استخدامًا في التعلم الآلي التعلم تحت الإشراف والتعلم غير الخاضع للرقابة.

1) التنقيب تحت الإشراف:

التنقيب تحت الإشراف هو عندما تقوم بتدريب نموذج باستخدام البيانات المصنفة .هذا يعني أن لديك بيانات لها بالفعل التصنيف الصحيح المرتبط بها .أحد الاستخدامات الشائعة للتعلم تحت الإشراف هو مساعدتك في توقع قيم البيانات الجديدة، ومع التنقيب تحت الإشراف، ستحتاج إلى إعادة بناء نماذجك أثناء حصولك على بيانات جديدة للتأكد من أن التوقعات التي تم إرجاعها لا تزال دقيقة .

من الأمثلة على التعلم تحت الإشراف وضع علامات على صور الطعام .يمكن أن يكون لديك مجموعة بيانات مخصصة فقط لصور البيتزا لتعليم نموذجك ما هي البيتزا[9].

2) التنقيب غير الخاضع للرقابة:

النقيب غير الخاضع للرقابة هو عندما تقوم بتدريب نموذج ببيانات غير مصنفة .وهذا يعني أنه سيتعين على النموذج العثور على ميزاته الخاصة وإجراء تنبؤات استنادًا إلى كيفية تصنيف البيانات.

مثال على التعلم غير الخاضع للرقابة هو إعطاء صور نموذجية لأنواع متعددة من الطعام بدون ملصقات .ستحتوي مجموعة البيانات على صور بيتزا، بطاطا مقلية، وأطعمة أخرى ويمكنك استخدام خوارزميات مختلفة للحصول على النموذج لتحديد صور البيتزا فقط دون أي ملصقات[10].

[9] (Palace, 1996)
[10] (dean، 2014)

7-1-3- تقنيات التنقيب في البيانات

هناك العديد من التقنيات والخوارزميات والطرق التي تتيح للمستفيدين القيام بالعديد من الوظائف باستخدام هذه التقنيات وإن اختيار التقنية المناسبة يعتمد على طبيعة البيانات التي مرت في المراحل السابقة وعلى حجمها:

7-1-3-1- التصنيف classification

التصنيف هو تقنية للتنبؤ بنتيجة معينة بناءً على مدخلات معينة. تستخرج نماذج تصف بشكل دقيق فئات وتصنيفات البيانات الهامة، وتتنبأ مثل هذه النماذج، بالتصنيفات الفئوية (المنفصلة وغير المرتبة).

تحلل الخوارزمية مجموعة تدريب تحتوي على مجموعة من السمات جنبًا إلى جنب مع نتائج كل منها (هدف أو سمة تنبؤ)، ويتم استخدام هذا بعد ذلك للتنبؤ بنتائج بيانات غير معروفة تسمى مجموعة الاختبار. أولاً، تحاول الخوارزمية اكتشاف العلاقات بين سمات مجموعة التدريب التي تجعل من الممكن التنبؤ بالنتيجة، وبعد ذلك يتم إعطاء الخوارزمية مجموعة بيانات لم يتم رؤيتها من قبل، تسمى مجموعة التنبؤ أو مجموعة الاختبار، والتي تحتوي على نفس مجموعة السمات، باستثناء سمة التنبؤ تكون (غير المعروفة) حتى الآن. [11]

تقوم الخوارزمية بتحليل المدخلات وإصدار التنبؤ. تحدد دقة التنبؤ مدى جودة "الخوارزمية". على سبيل المثال، يمكن استخدام التصنيف في قاعدة بيانات طبية للتنبؤ بما إذا كان المريض يعاني من أمراض القلب بناءً على المعلومات المسجلة سابقا والتي تحمل سمة الإخراج (يعاني/ لا يعاني)، مثال أخر تحتاج قروض البنك التحليل في بياناتها لمعرفة أي من المتقدمين للحصول على القروض ("آمنون" / "محفوف بالمخاطر") بالنسبة للبنك، أو توقع ما إذا كانت ستمطر في يوم معين أم لا.[12]

[11] (romero)
[12] (تشابل)

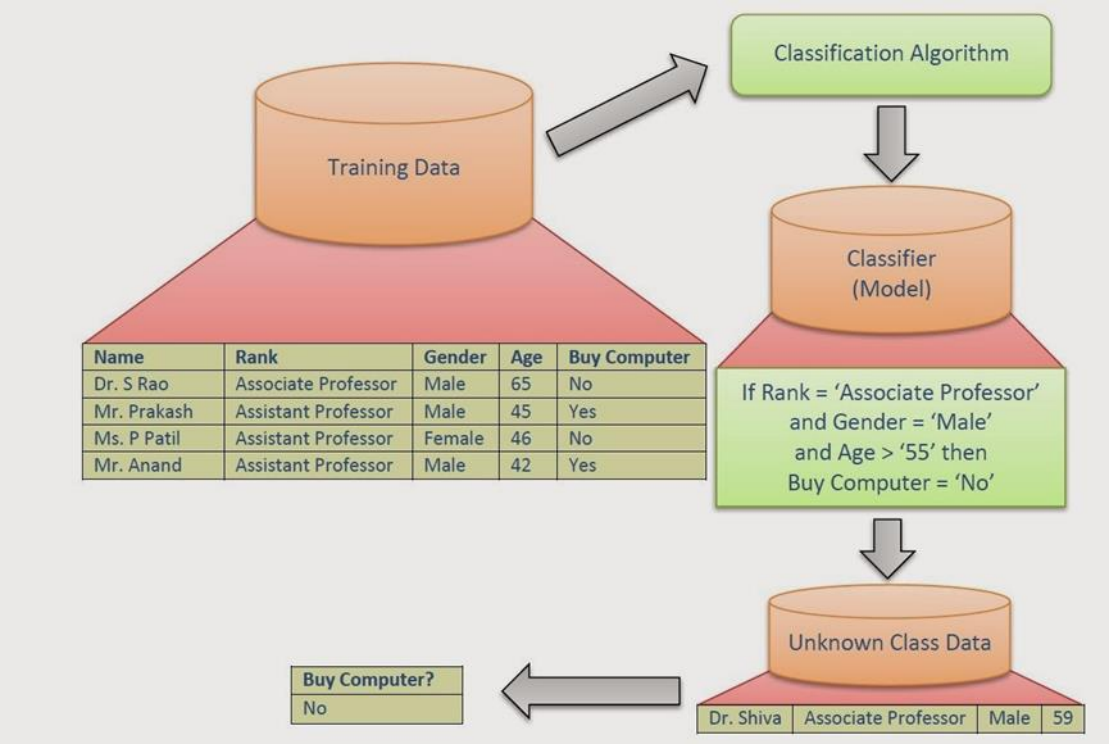
- **يتكون النهج العام للتصنيف بأنه عملية من خطوتين:**

1) الخطوة الأولى:

نقوم ببناء نموذج تصنيف بناءً على البيانات السابقة وتسمى خطوة التعلم أو مرحلة التعلم، وفي هذه الخطوة تقوم خوارزميات التصنيف ببناء المصنف. المصنف مبني على مجموعة التدريب (البيانات السابقة) المكونة من قاعدة بيانات tuples وتصنيفات الفصل المرتبطة (سمة التنبؤ).

2) في الخطوة الثانية:

نحدد ما إذا كانت دقة النموذج مقبولة، وإذا كان الأمر كذلك، فإننا نستخدم النموذج لتصنيف البيانات الجديدة، في هذه الخطوة يتم استخدام المصنف للتنبؤ ببيانات غير معروفة، هنا يتم استخدام بيانات الاختبار لتقدير الدقة قواعد التصنيف. يمكن تطبيق قواعد التصنيف على البيانات الجديدة، تتم مقارنة هذا المصنف مع المصنف الفعلي.[13]



الشكل (2)

[13] (romero)

- **العوامل المؤثرة على أداء دقة التصنيف:**

1) تحويل البيانات:Data transformation

تحتوي عملية تحويل البيانات علة عمليتين، وليس بل ضرورة استخدامهم معا أي أن طبيعة البيانات هي التي تحدد

أ- التطبيع: Normalization

يتم تحويل البيانات باستخدام التطبيع، ويشمل التطبيع قياس جميع القيم لسمة معينة لجعلها تقع ضمن نطاق صغير محدد. يستخدم التطبيع عندما يتم استخدام الشبكات العصبية أو الأساليب التي تنطوي على قياسات في خطوة التعلم.

ب- التعميم: Generalization

يمكن أيضًا تحويل البيانات من خلال تعميمها على المفهوم الأعلى لهذا الغرض يمكننا استخدام التسلسل الهرمي للمفاهيم.

2) وجود القيم المتطرفة: Presence of outliers

القيم المتطرفة هي نقاط بيانات لا تتوافق مع غالبية البيانات. القيم المتطرفة هي أيضًا نقاط يصعب تصنيفها بسبب يمكننا أن نقول انهم لا يملكون خصائص ناقلات ميزة مماثلة مثل غالبية البيانات. لذا تصنيف القيم المتطرفة مهمة محفوفة بالمخاطر ويمكن أن تؤثر على الدقة بشكل سلبي. طريقة جيدة يمكن استخدامها هنا مقدما هي إزالة القيم المتطرفة. هناك عدد قليل من الطرق الجيدة للقيام بذلك التجميع وتركيب المنحنى. [14]

3) إزالة ضوضاء: Noise removal

يمكن تعريف الضوضاء عادة على أنها خطأ عشوائي أو التباين في متغير مقاس يكون النوعان النموذجيان فيه غير متناسقين قيم الميزات أو الفئات. الضوضاء عادة ما تكون أقلية في مجموعة البيانات. ويمكن إزالته باستخدام خوارزميات العنقدة.[15]

[14] (Prediction)
[15] (romero)

4) تحليل مدى الارتباط: Relevance analysis

قد تحتوي البيانات في بعض الأحيان على مجموعات قليلة من الصفات التي لا توفر الكثير من المعلومات للمصنف. هذا يعني انه لا تشكل هذه السمات إدخالات مهمة في متجه الميزة. إزالة هذه السمات يمكن لهذه الصفات تسريع أداء المصنف. يمكن أيضا القيام بذلك باستخدام(Chi-square)، (LDA (Linear Discriminant Analysis تساعد هذه الطرق في تقليل مساحة الميزة بشكل فعال.[16]

5) طرق التقدير الخاطئة: Wrong estimation methods

لا ينبغي أن تكون دقة التصنيف مثالية تقاس في تجربة واحدة. التحقق المتقاطع هو طريقة جيدة جدًا للقياس الدقة التي تستخدم نوعًا من الإجراء وتترك الدقة لجميع التكرارات. لكن الدقة لا تزال تدبير شخصي ولا يمكن تزويدنا بمعلومات كاملة عن أداء المصنف.[17]

[16] (Prediction)
[17] (Prediction)

2-7-1-3- الخوارزميات التصنيف

1) خوارزمية (J48) لشجرة القرار: (j48 decision tree)

شجرة القرارات (Decision Tree)هي رسم تخطيطي، على شكل شجرة متفرعة يستخدم لتحديد مسار العمل أو تظهر ما هي الاحتمالات الممكنة. يمثل كل فرع من فروع شجرة قرار محتمل. يتم تنظيم الشجرة لإظهار كيف ولماذا قد يؤدي أحد الخيارات إلى الخيار التالي، مع استخدام الخطوط المتفرعة والتي تشير إلى أن كل خيار هو خيار مستقل عن الخيار الآخر، وتعد أشجار القرار أيضًا جزءًا أساسيًا من الغابات العشوائية [18]

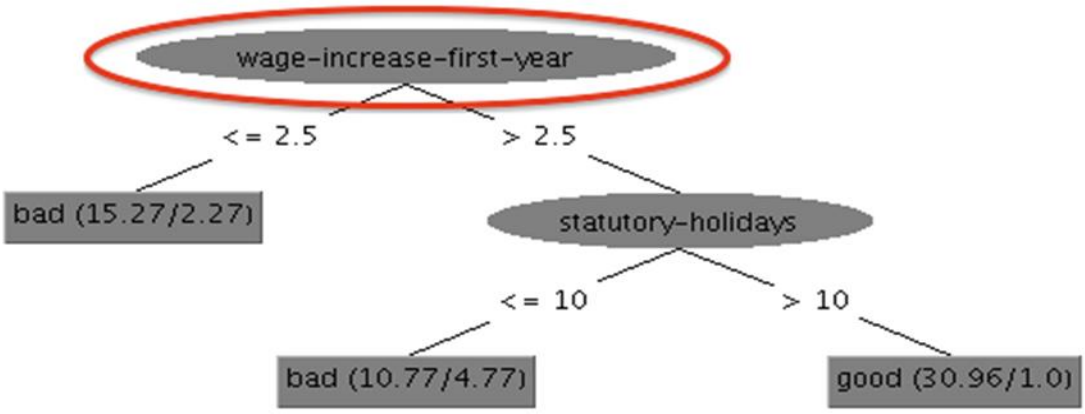
شجرة القرارات هي خوارزمية تعلم الآلة متعددة الوظائف، والتي يمكن أن تؤدي مهام التصنيف والانحدار، وحتى تشمل المهام متعددة المخرجات، وهي خوارزمية قوية جدًا يمكن أن تناسب مجموعات البيانات المعقدة للغاية.[19]

- **مصنف (J48): Classifier(J48)**

يندرج هذا المصنف ضمن خوارزميات أشجار القرار والتي على اختلاف أنواعها تشابه الى حد ما خوارزمية التصنيف (Naïve Bayes) من حيث اعتمادها على الاحتمالات الشرطية مع اختلاف رئيسي يكمن في هذه الخوارزمية تقوم بتوليد قواعد (Rules) لاستخدامها كجمل شرطية لتحديد السجلات والأحداث الاحتمالية بشكل عبارة شرطية (IF…. THEN).[20]

يستند هذا المصنف إلى هيكلة شجرية مؤلف من عقد رئيسية تدعى الجذر(Root) ومجموعة عقد داخلية (Nodes) ومجموعة عقد نهائية (Terminals) بحيث يتضمن كل الجذر ومجموعة العقد الداخلية القاعدة (Rule) التي تحدد المسار للفروع المرتبطة بما يسمح في النهاية بالوصول إلى النتيجة النهائية كما هو موضح بالشكل():[21]

[18] (Morgan & Teachman, nov/1988)
[19] (han, kamber, & pei, 2000)
[20] (سلق)
[21] (الضاهر، 2014)



الشكل (3)

تعتمد هذه الخوارزمية إلى تقسيم مجموعة بيانات التدريب المراد تصنيفها إلى مجالات متقاطعة (Mutual Exclusive) ذات تسمية أو قيمة أو عملية لتوضيح وشرح البيانات داخل هذا المجال وذلك بالاعتماد على معيار يستخدم لحساب أو تعين أفضل المعاير لتجزئة هذا المجال من البيانات التي يتم تدريبها والذي يدعى التابع الإحصائي (Information Gain) والمعروف بالمعادلة التالية:

$$Gain(S,A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

بحيث:

1) مجموعة بيانات التدريب (S) ومجموعة المعايير (A)
2) Values (A): جميع القيم الممكنة للمعيار A
3) (Sv) مجموعة جزئية من المجموعة (S) المنتمية للمعيار (A) وذات قيمة (V)
4) تابع العشوائية (Entropy): يعبر هذا التابع عن عشوائية المعطيات وتتراوح قيمته بين (0−1)

ويعبر عنه بالمعادلة التالية:

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

حيث يعبر المتغير (Pi) عن احتمالية انتماء مجموعة البيانات (S) إلى الفئة (i)

إن بناء الهيكلة لأغلب أشجار التصنيف يتم من الأعلى الى الأدنى (top–down) والاستفادة من طريقة البحث (Greedy) لتحديد قيم التابع (Gain) لمختلف المعاير واختيار المعاير ذات القيمة الأعلى للتابع (Gain) ومن ثم إعادة العملية لباقي المعاير وصولا لمجموعات جزئية متجانسة.

تتصف النماذج التي يولدها هذا النوع من التصنيفات بالدقة العالية والسرعة في بناء النموذج، كما يمكن تطبيقها على البيانات متعددة الفئات (Multi class) وبأنها قابلة للتأويل والفهم من خلال تحليل شجرة القرار ومعاينة الرسم البياني المولد عند بناء النموذج. [22]

[22] (سلق)

2) خوارزمية الغابات العشوائية: (Random Forests)

الغابة العشوائية هي خوارزمية تعلم خاضعة للإشراف وهي تطور لأشجار القرار تستخدم في كل من التصنيف والانحدار ومع ذلك، يتم استخدامه بشكل رئيسي لمشاكل التصنيف، كما نعلم أن الغابة تتكون في الواقع من العديد من أشجار القرار لتصنيف حالة جديدة يعني غابة أكثر قوة، وتمكن الغابة العشوائية عددًا كبيرًا من المصنفات الضعيفة أو المرتبطة بشكل ضعيف لتشكيل مصنف قوي.

تقوم خوارزمية الغابات العشوائية بإنشاء أشجار القرار على عينات البيانات حيث يتم تحديد مجموعة فرعية من الميزات بشكل عشوائي من الميزات الاختيارية لتنمية الشجرة في كل عقدة، وتزرع كل شجرة بدون تقليم بشكلها الأساسي، ثم تحصل على التنبؤ من كل منها، وفي النهاية تختار أفضل حل عن طريق تصويت جميع أشجار القرار إنها طريقة تجميع أفضل من شجرة قرار واحدة لأنها تقلل من التركيب الزائد وكلما زادت الشجرة كلما كانت النتيجة أكثر دقة.[23]

- **أنواع نماذج الغابات العشوائية :**

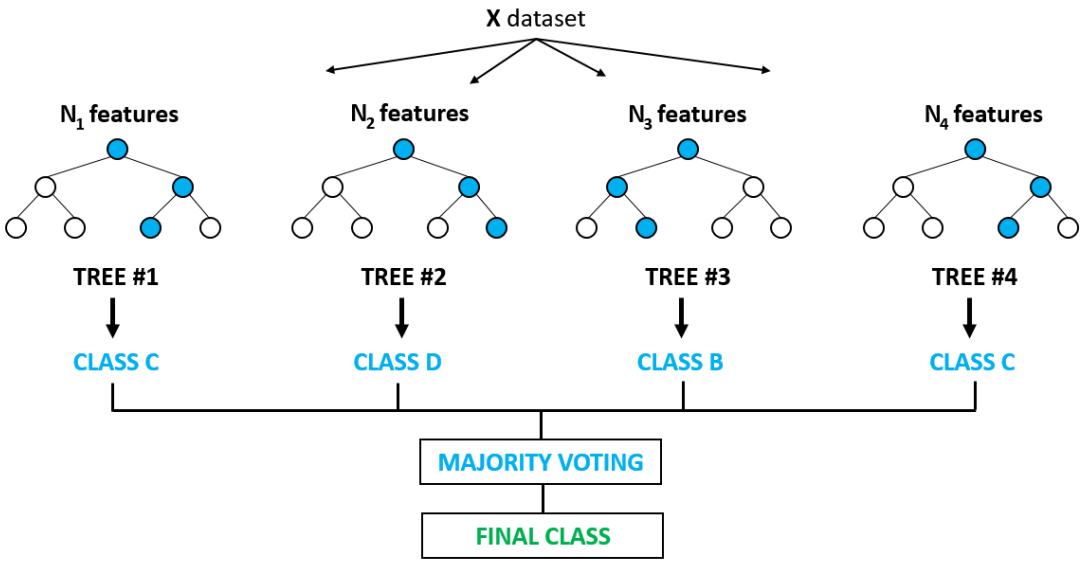
1) التنبؤ العشوائي للغابات **لمشكلة التصنيف:**

عند تصنيف كائن جديد وتمييزه بناءً على سمات معينة، ستعطي كل شجرة في الغابة العشوائية خيار التصنيف الخاص بها و "التصويت" وفقًا لذلك، وسيكون الناتج الكلي للغابة هو أكبر عدد من الأصوات، كما يوضح لنا الشكل (4) أن الناتج الكلي للغابة هو (C)[24]

2) التنبؤ العشوائي للغابات **لمشكلة الانحدار:**

خيار التصنيف في مشكلة الانحدار سيكون ناتج الغابة العشوائية هو متوسط جميع مخرجات شجرة القرار.[25]

[23] (Breiman, September 2010)
[24] (sejdinovic)
[25] (dean، 2014)


X dataset
N1 features
N2 features
N3 features
N4 features
TREE #1
TREE #2
TREE #3
TREE #4
CLASS C
CLASS D
CLASS B
CLASS C
MAJORITY VOTING
FINAL CLASS

الشكل (4)

3) مصنف بايز الاحتمالي: (Naïve Bayes Classifier)

نظرية بايز هي مفهوم مهم جدا في علوم التنقيب في البيانات، حيث لها استخدام واسع في التنقيب كمصنف classifier يستخدم تصنيفNaive Bayes

هي نظرية أساسية في علم الاحتمالات، حيث تحدد الاحتمال الشرطي لحدث ما. هذا الاحتمال الشرطي يعرف بالفرضية، ويتم حساب احتمالية حدوث هذه الفرضية بناء على معطيات أو دلائل سابقة، باختصار بإمكاننا تعريفها على أنها احتمالية حدوث حدث ما بمعلومية أن حدث آخر قد حدث بالفعل[26].

أما الشكل الرياضي للنظرية فيقدم كالآتي:

$$p(H \mid E) = \frac{p(E \mid H)\ p(H)}{p(E)}$$

مصنف بايز **(Naïve Bayes Classifier)** من أشهر أساليب تعلم الآلة، والتنقيب في البيانات والتصنيف على وجه التحديد clasification حيث يتميز بالسرعة في المعالجة والكفاءة في عمليات التنبؤ. يعتمد هذا الأسلوب على المفهوم الإحصائي لنظرية بايز الأساسية الذي ناقشناه بل أعلى والذي يحسب احتمالية حدوث نتيجة معينة بتحقق ما هو متاح ومعروف ويسمى ساذج، برئ(Naïve) لأنه يعتمد مبدأ Independence Assumptions بحيث ينظر للعلاقة بين جميع الخصائص Attributes, Features بأنها مستقلة عن بعضها البعض. [27]

بمعنى أن النموذج لا يعير اهتماما للعلاقة بين الخصائص إن وجدت فجميعهم يساهمون في حساب الاحتمال والنتيجة النهائية ستكون رقماً لا يحمل معنى من حيث توضيح اعتماد خاصية على أخرى أو قيمة للترتيب. كما يتميز نموذج بايز للتصنيف بسهولة البناء والتطوير والقدرة على معالجة البيانات الكبيرة ويتفوق في ذلك على عدد من الخوارزميات المتعقدة والمتقدمة. بحيث يتم تدريب النموذج بالبيانات وخصائصها المتاحة في قواعد البيانات ومن ثم

[26] (Qin، Xia، Wang، و Du)
[27] (han, kamber, & pei, 2000)

يقوم النموذج بتحديد نوع السجلات الجديدة وتصنيفها بالاعتماد على البيانات والإحصاءات المتوفرة سابقاً لديه، والشكل الرياضي للمصنف فيقدم كالآتي:[28]

$$P(x \mid c) = P(x_1 \mid c_j) \cdot P(x_2 \mid c_j) \cdot \ldots \cdot P(x_d \mid c_j)$$

$$= \prod_{k=1}^{d} P(x_k \mid c_j)$$

[28] (Raschka, October 4, 2014)

4) خوارزمية الجار الأقرب: (K–Nearest Neighbor KNN)

هي إحدى خوارزميات التنقيب في البيانات والتعلم الآلي والتي تعمل بمشرف (موجه) وتعد خوارزمية الجار الأقرب من خوارزميات التصنيف التنبؤية والوصفية ولديها الإمكانيات لتوليد التقديرات المحلية للنقطة، وتكون الطريقة سهلة في أغلب الأحيان عندما يكون حجم البيانات صغيرا، وتصف البيانات لمعرفة الجار الأقرب بسهولة وفعالة عندما تكون قاعدة البيانات كبيرة جدا ولها القدرة على التعامل مع البيانات (المشوشة) أي الناقصة، و تعتمد الخوارزمية في عملها على قياس المسافة الاقليدية بين كل نقطة والنقطة الأقرب إليها وعندما تكون البيانات قريبة من بعضها تكون المسافة الاقليدية قليلة جدا بين كل نقطة والنقطة المجاورة لها ولكن كلما تباعدت قيم البيانات وتبعثرت أصبحت المسافات بين النقاط كبيرة ومنها جاء عنوان الخوارزمية، إذ يشير الحرف K إلىُ الحالات التي سيتم تصنيفها بناء على المسافات بينها (أي بين الجيران)، حسب المسافة من المعادلة التالية:[29]

$$Euclidean\ Distance = d = \sqrt{\sum_{i=1}^{N} \left(Xi - Yi \right)^2}$$

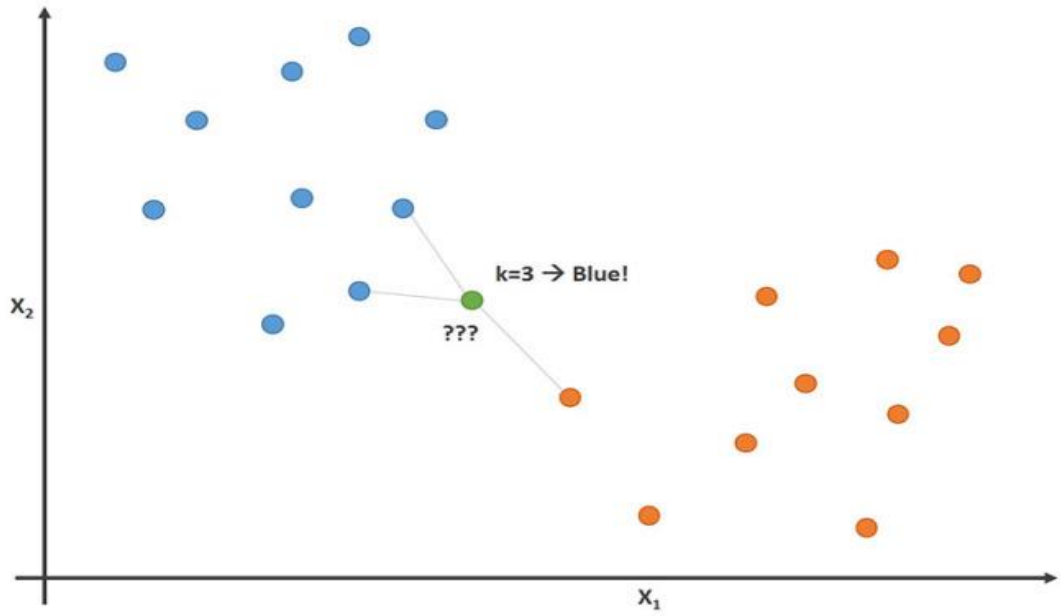
حيث أن: (d: المسافة بين أي نقطتين) و(Xj:Xi نقاط تموضع البيانات في البحث)

ولهذه الخوارزمية محددات مثلما لها مزايا، ومن أهم محدداتها انها تحتاج إلى وقت طويل عند العمل على البيانات الكبيرة، وكذلك فهي حساسة عند التعامل مع سمات التدريب ذات الخواص غير المترابطة، لأن هذا يضعف التصنيف ولهذا يجب اختبار سمات ذات الخواص المترابطة.[30]

- **خطوات عمل الخوارزمية:**
  - حدد عدد الجيران الأقرب ولتكن K
  - حساب المسافة الإقليدية بين السجل المستكشف وأقرب جار من المعادلة الرياضية السابقة[31]

[29] (الفخري، 2018)
[30] (Imandoust, Bafandeh, & Bolandraftar, 2013)
[31] (مطللك و عبد الحسن)

- ترتيب المسافات بإعطاء الرتب لها من أصغر مسافة الر أعلى مسافة، ثم حدد الجيران الأقرب بالاستناد الى مسافة حد أدنى (K−th)[32]



الشكل (5)

[32] (محمد و داود، 2012)

5) الانحدار اللوجستي: (Logistic regression)

هو أحد النماذج الاحصائية يستعمل لتحليل مجموعة من البيانات تحتوي على متغير واحد أو أكثر من المتغيرات المستقلة المراد معرفة تأثيرها في المتغير التابع، ويمكن أنُ يستخدم الانحدار الوجستي عدة متغيرات م تكون رقمية أو فئوية بالاعتماد على مجموعة متغيرات مستقلة ومختلطة، ويتميز المتغير التابع بأنه يحتوي على قيم ثنائية (نتيجتان محتملتان وحيدتان)

بمعنى آخر فان المتغير التابع يحتوي على قيم اسمية مثل (فشل، نجاح) أو (موت، حياة) وغيره، وهذا ما يسمى بالانحدار اللوجستي الثنائي (Binary Logistic Regression)

ويتميز الانحدار اللوجستي بأنه النموذج الاكثر مرونة من نماذج الانحدار التقليدي اذ انه لا يشترط ان تكون العلاقة بين المتغيرات المستقلة والمتغير التابع خطية، وايضا لا يشترط بأن تكون المتغيرات التوضيحية من النوع المستمر او المتقطع ولا ان يكون لها التوزيع الطبيعي، ويتميز أيضا بأن العلاقة بين المتغير التابع والمتغير المستقل غير خطية التي يمكن تحويلها الى الشكل الخطي باستعمال ما يعرف بتحويلة اللوجت (LOGit tarnsformation).[33]

يعتمد تصنيف الانحدار اللوجستي على البيانات الموجودة لإنشاء صيغة انحدار لحدود التصنيف (البحث عن الأنسب)، وذلك للتصنيف. الانحدار هنا يعني العثور على أفضل مجموعة معلمات ملائمة ومجموعة معلمات دالة متعددة المتغيرات والانحدار غير الخطي، تتم طريقة تدريب المصنف في الانحدار اللوجستي هي العثور على أفضل المعلمات المناسبة، ويتم استخدام خوارزمية التحسين. تختلف عن خوارزمية التصنيف السابقة، فهي فقط لضبط حجم المعلمة من أجل العثور على الأمثل. [34]

[33] (Morgan & Teachman, nov/1988)
[34] (النصراوي، 2017)

- **أنواع الانحدار اللوجستي:**

1) **الانحدار اللوجستي الثنائي** (Binary Logistic Regression) وهذا النوع يعد الأكثر شيوعا، ويستخدم هذا النوع عنما يكون المتغير التابع يحتوي على قيم اسمية مثل (فشل، نجاح) أو (موت، حياة).

2) **الانحدار اللوجستي المتعدد** (Multinomial Logistic Regression) ويستخدم هذا النوع عنما يكون المتغير التابع متغيرا اسميا متعدد الاستجابة.

3) **الانحدار اللوجستي الترتيبي** (Ordinal Logistic Regression) ويستخدم هذا النوع عنما يكون المتغير التابع رتبويا[35].

[35] (Morgan & Teachman, nov/1988)

الفصل الرابع

الإطار النظري

سنتكلم في هذا الفصل عن مفهوم التسويق البنكي وأهميته، وعن مفهوم التسويق المباشر، والودائع البنكية وأنواعها بشكل عام، وحاجة البنوك للتنقيب في بياناتها وسنتحدث عنهم كالتالي:

التسويق البنكي

(مفهوم، أهداف، وظائف إدارة التسويق، أهمية، خصائص)

التسويق المباشر

(مفهوم، ميزات وعيوب، أنواع)

الودائع البنكية

(مفهوم، أنواع)

(حاجة البنوك للتنقيب في البيانات)

تمهيد:

يعتبر التسويق من أهم الوظائف الأساسية بالنسبة لأي مؤسسة، وأحد أدواتها الديناميكية في تحقيق أهدافها، لدوره الحيوي في ضمان البقاء والاستمرار، الأمر الذي جعل جل المؤسسات الخدمية خاصة منها المصرفية تدرك هذا الدور وتعمل على الاهتمام به أكثر فأكثر، نظرا لما يوليه من اهتمام بدراسة السوق التي تقوم على معرفة حقيقية لواقع الزبائن الحاليين والمرتقبين، وكذلك دراسة حاجاتهم لأجل كسب رضائهم وولائهم بما يحقق للمؤسسة المصرفية أهدافها ويضمن بقائها واستمدادها، وهكذا كان التسويق المصرفي ردا استراتيجيا من جانب هذه الأخيرة، استجابت من خلاله لمقتضيات التغيرات الحاصلة على الساحة المصرفية .وسيتم في هذا الفصل تكوين ما يمكن اعتباره مدخلا للتعرف على مفهوم التسويق المصرفي والمضامين الأساسية التي ينطوي عليها على النحو التالي.

1-4- التسويق المصرفي

1-1-4- مفهوم التسويق المصرفي

يختلف مفهوم التسويق المصرفي من مفكر الى آخر بسبب اختلاف زوايا نظرهم للتسويق المصرفي فسوف نتناول عدة مفاهيم:

- ✓ عرف موخ **(Mauch)** التسويق المصرفي انه الطريقة التي تستطيع من خلالها المصارف تحقيق اهدافها وتلبية حاجات السوق وتحويل الحاجات الى طلب حقيقي
- ✓ اما واير **(Wayer)** فقد اعتبر التسويق المصرفي نشاط ينطوي على تحديد أكثر للأسواق ربحية في زمن الحاضر والمستقبل، بالإضافة الى تقييم الحاجات الحالية والمستقبلية للعملاء، فهو يتعلق بوضع اهداف المصرف وتصميم البرامج اللازمة لتحقيق تلك الأهداف[36].
- ✓ في حين يرى الاقتصادي **(Hodages Tillman)** أن التسويق المصرفي هوٌ الاستمرار في تقديم خدمات مصرفية متميزة وبكفاءة عالية بهدف رضاء العميل وإظهار صورة متميزة للبنك مع تحقيق أقصى الأرباح للبنك.
- ✓ التسويق المصرفي الاقتصادي **(P.Kotler)** الذي أعتبر التسويق المصرفي على أنه مٌجموعة الأنشطة المتكاملة التي تجرى في إطار إداري محدد، وتقوم على توجيه انسياب خدمات البنك بكفاءة وملائمة لتحقيق الإشباع للمتعاملين من خلال عملية مبادلة تحقق أهداف البنك، وذلك في حدود توجهات المجتمع.
- ✓ وهناك تعريف أخر ينص على ان التسويق المصرفي هو تخطيط منظم ومتواصل ومتكامل الدراسة الحاجات المالية المتغيرة التي تقدمها المؤسسة المصرفية لإشباع حاجات المستفيدين من خلال برامج فعالة.

خلاصة عامة يمكن القول ان التسويق المصرفي هو مجموعة من الانشطة المتخصصة والمتكاملة التي توجه من خلال موارد المصرف وامكانياته، ويستهدف تحقيق مستويات اعلى من إشباع الحاجات ورغبات العملاء الحالية والمستقبلية والتي تشكل دائما فرصا سوقية سانحة تستطيع من خلالها المصارف تحقيق الربح.[37]

---

[36] (بن ذيب و نصر الله، 2016)
[37] (فجخي، 2017)

2-1-4- لأهداف التسويق المصرفي

يهدف التسويق المصرفي إلى ابتكار وأداء الخدمات وإيصالها نحو الزبائن عن طريق منافذ مناسبة باستخدام وسائل الترويج المتاحة، وهذا قصد إرضاء الزبائن الحاليين وجذب زبائن جدد بما يحقق ا لربحية للبنوك، وبذلك يساهم التسويق المصرفي في تحقيق جملة من الأهداف تتمثل فيما يلي:[38]

1) تحسين سمعة البنك وذلك ببناء صورة ذهنية ايجابية عنه وعن خدماته وعن العاملين فيه.
4) ومعرفة مدى قدرة على التأثير في السوق.
2) تكييف البنوك وفقا لمتغيرات تحقيق الأهداف المالية المتمثلة في أهداف السيولة، الربحية والأمان مع تحقيق نمو موارد البنك.
3) أهداف توظيف الأموال، وهذا عن طريق:
   أ- زيادة حجم القروض والسلفيات.
   ب- زيادة الاستثمارات في الأوراق المالية.
5) ابتكار خدمات مصرفية جديدة تستجيب لرغبات الزبائن، والعمل على تطوير الخدمات المصرفية الموجودة.
6) متابعة البنوك المنافسة السوق وتطورات احتياجات الزبائن.
7) خلق أسواق مصرفية جديدة عن طريق اكتشاف أنواع جديدة من الخدمات المصرفية التي يرغب فيها الزبون.
8) دراسة الانتشار الجغرافي لوكالات البنك في السوق المصرفية، والتخطيط لفتح وكالات مصرفية جديدة.
9) اكتشاف الفرص الاستثمارية، ودراستها وتحديد المشاريع الجديدة التي يمكن إقامتها.

[38] (بن ذيب و نصر الله، 2016)

3-1-4-وظائف إدارة التسويق المصرفي

تعتبر إدارة التسويق المحرك الأساسي للنشاط البنكي، وبموجب هذه الإدارة يتم مراقبة ومتابعة الانطباع الأولي للعميل عن تشكيلة الخدمات المصرفية المقدمة، والمستمدة أساسا من السوق المالية والمصرفية، وبذلك تعد الجهة المسؤولة عن إدارة الأنشطة التسويقية، ووضع الخطط، وسبل تنفيذها، والتأكد من تحقيق الأهداف المنشودة من كل ذلك، وإلى جانب ذلك تتولى مجموعة من الوظائف تتمثل فيما يلي:[39][40]

1) التعرف على احتياجات العملاء ورغباتهم.
2) القيام ببحوث تسويقية دقيقة لدراسة متطلبات السوق الحالية والتنبؤ بالاحتياجات المستقبلية.
3) دراسة عناصر المزيج التسويقي من خلال:

- إعداد وتطوير مزيج المنتجات المصرفية لمقابلة طلبات العملاء، ومنه دراسة ما تقدمه البنوك المنافسة من خدمات.
- اختيار الأساليب الترويجية والإعلانية المناسبة قصد الإعلان عن نشاط البنك والترويج لمنتجاته المصرفية لجذب العملاء إليه.
- فتح قنوات توزيع من فروع البنك لتقديم المنتجات المصرفية، ودراسة الانتشار الجغرافي لفروع البنك في السوق المصرفية وإمكانية فتح وحدات جديدة.

4) تقييم النشاط التسويقي للبنك مقارنة بالبنوك المنافسة خلال فترات متقاربة لتحديد نقاط الضعف والقوة لإدارة البنك العليا، ومن الأفضل أن تتبع إدارة التسويق مباشرة لإدارة العامة للتسويق في البنك والتي تتكون من: إدارة بحوث السوق والمعلومات، إدارة تخطيط الخدمات المصرفية، إدارة الترويج والبيع الشخصي، إدارة متابعة النشاط التسويقي على مستوى الفروع.

[39] (بن ذيب و نصر الله، 2016)
[40] (البغى، 2019)

4-1-4-أهمية التسويق المصرفي

يتكون النشاط التسويقي من مجموعة من الأنشطة التي تشمل عمليات التخطيط والتوجيه ورقابة الأداء بهدف تطوير الوضع التنافسي للبنك داخل السوق المصرفية، ويتم ذلك من خلال العمل على الحفاظ على العملاء الحاليين وتوسيع دائرة العملاء عن طريق الخدمات المصرفية بشكل متطور ومتميز، وقد ظهرت الحاجة لعملية التسويق نتيجة لازدياد حدة المنافسة بين البنوك التجارية من جهة والمؤسسات المالية الأخرى من جهة أخرى، بالإضافة إلى التغيرات السريعة في المؤثرات الاقتصادية والاجتماعية التي تحكم نشاط هذه المؤسسات والتي بدورها أوجبت على البنوك العمل على مواكبة هذه التغيرات للمحافظة على مركزها التنافسي.[41]

4-1-5-خصائص التسويق المصرفي

تتحدد خصائص التسويق البنكي في النقاط التالية:

1) المصرف وزبائنه (من موردين وهم مدخرين، وزبائن وهم مستثمرين) يشكلون سوقا تتميز بصفات خاصة بفضل العلاقة التي تربطهم فيما بينهم.

2) وجود تعامل مباشر مع العميل واتصال دائم مع المصرفي واحتكار كل بنك لشبكة خاصة به.

3) وجود أهمية كبيرة في تقسيم السوق المصرفية إلى مؤسسة كبيرة، ومتوسطة وأخرى صغيرة والعلاقة مزدوجة بين البنك والسوق، والتفاعل دائم بين سوق الاستعمالات وسوق الموارد.

4) يتميز التسويق في البنوك بالدور المزدوج بجذب الودائع والمدخرات كمادة أولية من جهة، ومنح القروض وعرض المنتجات البنكية من جهة أخرى.

5) نقاط بيع الخدمات البنكية هي أماكن إنتاجها، وهي نفسها أماكن التوزيع المتمثلة في الوكالات.

[41] (الحسون، 2016)

6) تؤثر اللوائح والقوانين والتشريعات الحكومية على المؤسسات المالية والبنكية في تقديم عروضها وتحديد أسعارها، وتكون المنافسة غير كاملة لوجود قوانين تحد من القدرات المختلفة.

7) تكون الدورة التوزيعية للخدمات البنكية وحدوية، أي من المدخر إلى البنك إلى المستثمر.

8) محاولة التجسيد المادي لخدمات البنوك من خلال الإشهار، الاستماع، الاستشارة.

وعموما تتمحور خصائص التسويق البنكي حول فكرتين رئيسيتين:

1) تشابك العلاقات التي تربط البنك بمكونات المحيط المختلفة.
2) تعدد علاقات البنك بالمحيط ومكوناته المختلفة.

2-4-التسويق المباشر

4-2-1- مقدمة

أسهم تطور الخدمات المالية عموما والمصرفية بالخصوص في نمو وازدهار أساليب التسويق المباشر، وأصبح شعار وأصبح شعار البنوك هو "وجود مجتمع بدون نقود"، فعلى سبيل المثال يعد مصرف (Générale Société) من المصارف الرائدة في استعمال أدوات التسويق المباشر مثل إطلاق خدمة (Vocalia) و (Logitel) من أجل السماح للعملاء بالقيام بمختلف العمليات المصرفية عن طريق الهاتف مثل الاطلاع على الرصيد، وتحويل الأموال والتواصل مع مستشار خاص، واستهداف العملاء من خلال المكالمات الهاتفية إن كانت المكالمة خارجة من البنك (خارجية) أو من العميل أثناء استفساره (داخلة)، والرسائل، ويتم من خلال المكالمة تسجيل بيانات هذا العميل الجديدة أو تعديلها لتحسين الحمالات التسويقية بعد عمليات القياس والتحليل.

4-2-2- مفهوم التسويق المباشر

يعرف التسويق المباشر بأنه أحد أنواع الحملات الإعلانية والذي يعمل على استخدام القنوات الاتصالية المختلفة مثل (البريد الإلكتروني، الاتصال بالهاتف، طلب زيارة المتجر وغيرها)، فهو في الأساس يهدف لجعل العميل يأخذ خطوة فعلية تجاه ما يقدمه المعلن وهو ما يسمى ب (Call To Action)، قد يتطلب الأمر بعض الوقت ولكن في النهاية يحقق الهدف المطلوب، ويتميز ذلك النوع بالقدرة على قياس النتائج بشكل أسرع وأفضل. كما أنه يعمل على تحديد شريحة محددة ودقيقة من الجمهور المستهدف.[42]

4-2-3- مميزات وعيوب التسويق المباشر

(لكل شئ مميزات وعيوب، سوف تتعرف الآن عليها، ولنبدأ بالمميزات وهي كالآتي):

- ✓ **الرسالة الإعلانية تحديد الجمهور المستهدف:**

فذلك النوع من الحملات الإعلانية يجعلك قادراً على تحديد الجمهور المستهدف الذي ترغب في توصيل إلي، قد يتخذ الأمر بعض الوقت في تحديد الشرائح المستهدفة ولكن في النهاية

[42] (جسير، 2018)

سوف تصل إلى المستهلكين المحددين وترسل إليهم رسالتك الإعلانية التي تود أن يتفاعلوا معها.

- ✓ **الاحتفاظ بالجمهور السابق والتواصل مع جمهور جديد:**

يتيح ذلك النوع من الحملات بأنك تعمل على التواصل المستمر مع جمهور القديم، عن طريق إرسال الرسائل البريدية أو الاتصال بهم وتطلعهم على كل جديد، لكن من الممكن أيضاً أن تقوم بتحديد شرائح مستهدفة جديدة لتنضم إلى قائمة عملائك وترسل لهم ما تقوم ببيعه والتواصل معهم بشكل مستمر، بالتالي فأنت ستضمن على بقاء عملائك السابقين، ووجود عملاء حاليين مما يترتب عليه زيادة نسبة المبيعات لديك.

- ✓ **معرفة متطلبات الجمهور:**

يتميز التسويق المباشر بأنك سوف تتابع جميع تطورات سوق العمل متطلبات الجمهور، سواء عن طريق الاتصال المباشر أو انتظار الرد عبر الرسائل البريدية.

- ✓ **القدرة على قياس النتائج:**

سوف يتيح لك أيضاً الفرصة لمتابعة وقياس النتائج بشكل أسرع ومعرفة ما يجب تحسينه وتطويره.

- ✓ **تحديد التكلفة:**

يتميز ذلك النوع أيضاً بأنك تستطيع أن تحدد التكلفة الخاصة به، كما يعمل على تقليل الوقت والجهد إذا كنت سوف تتواصل عبر البريد الإلكتروني، وذلك لأنك سوف ترسل رسالة مطابقة لجميع العملاء المستهدفين.

تلك تعد أبرز المميزات الخاصة بالتسويق المباشر، ولكن هناك عيب واحد وقد يكون لا يعد مشكلة كبيرة وهو أنك تقوم بتحديد الجمهور المستهدف بشكل محدد ودقيق. [43]

[43] (مسالتة، 2017)

4-2-4- أنواع التسويق المباشر

- ✓ **البريد الإلكتروني:**

واحد من الطرق السريعة، السهلة وتتميز بمحدودية التكلفة، فهو يعمل على إرسال الرسائل البريدية والإعلانات، النشرات البريدية وآخر التطورات لمجموعة من العملاء السابقين والمرتقبين في آن واحد.

- ✓ **الرسائل النصية(SMS)**

أحد الأنواع المنتشرة، يتم عن طريق جمع البيانات الخاصة بالعملاء وأبرزها رقم الهاتف والاسم، ثم إرسال الرسائل إليهم للتنويه عن عروض أو تخفيضات.

- ✓ **وسائل التواصل الاجتماعي:**

يمكنك التواصل مع العملاء عن طريق وسائل التواصل الاجتماعي وتقديم العروض والخدمات الخاصة لهم، كما أصبح هناك خاصية إرسال الرسائل عبر منصة" فيس بوك " والتي تقدم فيها أبرز الخدمات والعروض الخاصة بمنتجك.

والتسويق المباشر عبر الهاتف هو محور هذا البحث لما له مزايا عن باقي الأنواع:

- ✓ **التسويق عبر الهاتف:**

ذلك النوع يتم عقب تحديد الجمهور المستهدف بشدة وجمع البيانات الخاصة بهم، ثم يتم التواصل مع العملاء المرتقبين والعمل على تقديم الخدمات والمنتجات إليهم.

فهي وسيلة يمكنك من خلالها البيع الفوري مع العميل وتسجيل البيانات الخاصة به، كما أنك من خلالها يمكنك أن تتعرف على آراء العملاء ومتطلباتهم[44].

[44] (مسالتة، 2017)

3-4-الودائع البنكية

1-3-4- مفهوم الوديعة

الوديعة البنكية هي مبلغ من المال يتم حفظه أو إيداعه في البنك لمدة معينة محددة أو غير محددة وفق عقد وشروط مبرمة يتفق عليها الطرفين صاحب المبلغ والبنك، ويكون ذلك بهدف حفظ المال أو توفيره أو استثماره، وبالتالي يحصل صاحب المبلغ على عائد محدد وفق شروط العقد طيلة وجود المبلغ النقدي في البنك[45]

2-3-4- أنواع الودائع

- **الودائع الجارية:** هي تلك الودائع التي يستطيع أصحابها سحبها دون سابق إنذار ولهذا السبب فان معظم البنوك لا تدفع فوائد على هذا النوع من الودائع.
- **ودائع التوفير:** يفتح هذا النوع من الحسابات لتشجيع الأفراد على الادخار مقابل فائدة معينة.
- **الودائع بإخطار:** أو ما تعرف بودائع تحت الطلب، وفيها يخطر المودع بنكه بالتاريخ الذي فيه يسحب وديعته.
- **الودائع لأجل:** هي تلك الودائع التي لا يستطيع أصحابها سحب منها إلا بعد المدة الزمنية المحددة المتفق عليها بين البنك والمودع ويتقاضى هذا الأخير فائدة على ذلك حسب المبلغ المودع ومدته.

وسيشمل هذا البحث الودائع الاستثمارية بشكل عام وغاليا ما تكون هذه الودائع هي الأشهر من بين كافة أشكال الودائع التي عادة ما تتيحها البنوك، لأنها تعد شكل من أشكال الاستثمار الغير مباشر بالنسبة لصاحب المال، وتتعرف بالتالي:[46]

الودائع الاستثمارية: هي مبلغ من المال الذي يتم إيداعه في البنك من قبل الأفراد أو المؤسسات على أن يتعهد البنك بحفظ المبلغ وإرجاعه عند الطلب أو بعد مدة زمنية محددة يتفق عليها بشروط العقد، طبعًا مضافًا إليها الأرباح والعوائد التي تم الاتفاق عليهًا أيضًا في العقد بين الطرفين.

---

[45] (الاخضر، 2013)

[46] (يوسف، يوسف سعادة، و محمد الزيدانيين، 2016)

4-4-حاجة البنوك للتنقيب في البيانات

يمكن للبنوك أن تستخدم البيانات التي تجمعها من خلال خدماتها التسويقية الهاتفية للتنبؤ برضى العميل وبنجاح الحملات التسويقية المباشرة المتبعة مع العملاء وتصنيف العملاء وفرزهم ضمن مجموعات لتسهيل عملية استهداف الشرائح، والخدمات الإلكترونية للتنبؤ بمخاطر الاحتيال والمصادر المحتملة له وبالتالي يمكن توفير الحلول المناسبة للتصدي له. كما يمكنها تحليل مشاعر العملاء سواء من خلال تحليل البيانات التي يتم جمعها من خلال تقييم العملاء المباشر للخدمات والمنتجات أو من خلال تحليل البيانات المتوفرة من خلال أطراف أخرى مثل وسائل التواصل الاجتماعي أو مواقع التقييم والتصنيف المالية. مما قد يكون له أثر كبير في الوصول الى رضا العميل.

بالإضافة الى ذلك، يمكن تحليل بيانات العملاء للوصول الى فهمهم بشكل أكبر حيث يمكن استهدافهم بالخدمات والمنتجات التمويلية المناسبة التي تلبي احتياجاتهم وتشبع رغباتهم. كما يمكن استخدام تلك البيانات لتحديد الجمهور المناسب لعمليات التسويق. حيث يمكن تحقيق ذلك من خلال تقسيم العملاء الى شرائح معينة بناءً على الأصول التي يمتلكونها او معاملاتهم أو العمليات المالية التي يقومون بها، فمثلاً يمكن تصنيفهم بناء على اصولهم من العقارات، الاستثمار في العملات أو الأسهم، حيث يتم استهداف العملاء بالحملات الترويجية والتسويقية وفقًا لشرائحهم وكذلك يمكن تصميم منتجات وخدمات تلبي رغباتهم. كذلك يمكن الذهاب ابعد من ذلك في مجال التسويق من خلال استخدام التسويق الشخصي بدلاً من التسويق القائم على الشريحة. حيث يقوم التسويق الشخصي على فهم الاحتياجات والعادات الشرائية الفردية للعملاء ومن ثم استهدافهم بالمنتجات والخدمات المناسبة لهم.

كما يمكن توظيف البيانات الضخمة في إدارة المخاطر، حيث يمكن أن تساهم في تحسين القدرة التنبؤية لنماذج المخاطر وزيادة دقتها وموثوقيتها وهذا بدوره سوف يساهم في توفير كبير في التكاليف. بالتالي في ضل نمو الطلب وتزايد التعقيدات، القدرة على والوصول إلى البيانات الضخمة واستخدامها سيلعب دوراً هاماً بنجاح إدارة المخاطر[47].

[47] (Veldhoen & De Prins, 2004)

الفصل الخامس

سيتم تقسيم هذا الفصل إلى مبحثين:

المبحث الأول: سيتحدث بلمحة سريعة عن البنك الذي أجريت عليه دراسة هذا البحث، ثم سنتحدث عن الأدوات المستخدمة بمجال التنقيب عن البيانات والتي سوف نحتاج اليها في هذا البحث، ومن ثم سيتم تقديم محتوى البيانات التابعة للبنك وتفسيرها وشرحها.

المبحث الثاني: فسيتحدث عن تصوير البيانات (Data Visualize)، وتجهيز البيانات وحل مشكلة عدم التوازن من أجل استخدام تقنيات التنقيب في التصنيف (Classification) ومن ثم اختيار المصنف الأفضل

1-5- المبحث الأول

1-1-5- لمحة عن البنك (Banco Português de Investimento)

يعود أصل (BPI) إلى عام 1981 مع إنشاء (SPI) (Sociedade Portuguesa de Investimentos)، مع هيكل مساهم متنوع، والذي تضمن مكونًا قويًا من نسيج الأعمال الوطني، يتكون من 100 من أكثر الشركات البرتغالية ديناميكية وأربع من أهم المؤسسات المالية المعايير الدولية.

في عام 1985، تم تحويل (SPI) وأدى إلى ظهور (BPI) ، أول بنك خاص تم إنشاؤه بعد إعادة فتح القطاع للقطاع الخاص بعد تأميم عام 1975. في عام 1986، أصبح BPI أول بنك مدرج في البورصة البرتغالية.

في عام 1991، بعد عقد من إنشائها، قامت (BPI)، التي حققت بالفعل قيادة واضحة في المجالات الرئيسية للخدمات المصرفية الاستثمارية، بتوسيع نشاطها إلى الخدمات المصرفية التجارية، من خلال الاستحواذ على.(BFB)

بين عامي 1996 و2005، واصل البنك نموه من خلال عمليات الاستحواذ والاندماج مع البنوك الأخرى، وفي عام 2002 تم إنشاء (Banco de Fomento) في أنغولا.

في عام 2006، أكملت 25 (BPI) عامًا من النشاط وواصلت استراتيجيتها المستمرة لخلق قيمة للمساهمين والموظفين والعملاء.

في عام 2012، نفذت BPI خطة إعادة الرسملة التي تضمنت إصدار 1.5 ألف مليون يورو من سندات التحويل الطارئة الثانوية (CoCo) التي اكتتبتها الدولة البرتغالية، من أجل الامتثال لعملية إعادة الرسملة التي اقترحتها .EBA في عام 2014، قام Banco BPI بسداد تلك الصكوك بالكامل، واختتمت عملية السداد للدولة قبل ثلاث سنوات من نهاية المدة المقررة.

في أبريل 2016، قامCaixaBank ، SA، الذي يملك 44.1٪ من رأس مال بنك Banco (BPI)في ذلك التاريخ، بنشر الإعلان الأولي عن عرض اكتساب عام وعام وطوعي (عرض) على جميع الأسهم التي تمثل رأس المال Banco BPI SA .بسعر 1،113 يورو للسهم . [48]

[48] (https://www.bancobpi.pt/grupo-bpi/grupo-bpi-grupo-bpi, n.d.)

في عام 2017، أكمل (Banco BPI) عملية نقل ملكية لصالح Unitel، تمثل 2٪ من رأس المال وحقوق التصويت لبنك .Banco de Fomento Angola نتيجة لإتمام هذا التحويل، أصبحت حيازات Banco BPI و Unitel في48.1 BFA ٪ و 51.9٪ على التوالي.

بعد فقدان حالة الشركة المملوكة ملكية عامة المعتمدة في الاجتماع العام في يونيو 2018 والاستحواذ القوي الذي أطلقه المساهم CaixaBank، SA استحوذ الأخير في 27 ديسمبر 2018 على رأس مال بنك Banco BPI بالكامل.

2-1-1-5- هوية البنك

تتميز هوية Banco BPI بالثقافة المالية والتجارية لبنك (Banco Português de Investimento) السمات الأساسية لهذه الثقافة هي استقلالية الإدارة والمرونة التنظيمية والعمل الجماعي وتمييز الجدارة والقدرة على التوقع والإدارة الصارمة للمخاطر والخلق الآمن للقيمة.

إن الربحية الكافية للبنك، من خلال أفضل ممارسات الإدارة والخدمة، هي هدف أساسي لنشاطنا حماية مصالح العملاء، مع التفاني والولاء والسرية، هي واحدة من المبادئ الأولى لأخلاقيات العمل وقواعد السلوك لموظفي البنك.

3-1-5- العلامة التجارية

السمات الأساسية ل (Banco BPI الخبرة والانسجام) موجودة في الألوان والطباعة ورمز علامتنا التجارية لذلك، فإن العلامة التجارية هي أكثر من مجرد صورة: إنها إعلان مبادئ والتزام تجاه العميل، أول سبب لدينا[49].

الشكل (11)

[49] (https://www.bancobpi.pt/grupo-bpi/grupo-bpi-grupo-bpi, n.d.)

2-1-5- الأدوات المستخدمة في التنقيب عن البيانات

1-2-1-5- أداة (weka)

(WEKA) برنامج مفتوح المصدر يوفر أدوات للمعالجة المسبقة للبيانات وتنفيذ العديد من خوارزميات التعلم الآلي وأدوات التصور حتى تتمكن من تطوير تقنيات التعلم الآلي وتطبيقها على مشاكل استخراج البيانات في العالم الحقيقي .يتم تلخيص ما يقدمه البرنامج في الرسم التخطيطي التالي فهو مجموعة من الأدوات والبرمجيات المتقدمة التي تساعد المستخدم في التنقيب عن البيانات وتجميعها وتحليلها باحترافية من خلال خوارزميات معقدة كتعلم الآلة والتعدين وغيرها وكل هذا يساعد على معالجة البيانات بشكل قوي وتجربة القواعد والخوارزميات عليها وحتى يمكن تحرير البيانات وتغيير سماتها وتصنيفها دون استهلاك الكثير من موارد جهاز الكمبيوتر، واختصاره (Waikato Environment for Knowledge Analysis)

وحزمة ويكا (Weka) توفر للمستخدمين مجموعة من الأدوات ومخططات التعلم (خوارزميات متقدمة) والتي تساعد على التنقيب عن البيانات واستخراجها بكل سهولة ويمكن الاستعانة بتلك الخوارزميات وتطبيقها بشكل مباشر على مجموعة من البيانات أو استخدامها بواسطة الجافا كود الخاصة بك بدون مشكلة، حزمة (Weka) تضم 4 أدوات متاحة ويمكن الوصول إليها وهم (Explorer, Experimenter، Knowledge Flow) و(Simple CLI)، وتسمح لك تلك الأدوات المهمة فتح مجموعة من البيانات وتحريرها كما يحلو لك بكل سلاسة، كما يمكن تغيير محتويات البيانات وتغيير السمات وتصنيف البيانات المتاحة وفقا لمجموعة محددة مسبقا من القواعد.[50]

علاوة على إمكانية إجراء تحليل التكلفة والمنفعة واستعراض مصفوفة التكلفة ومنحني العتبات، وعبر الأدوات التي يوفرها برنامج ((Wekaيمكن تجميع البيانات واستخدام الخوارزميات والقواعد وخصائص التقييم لتخطيط البيانات وعرض وتحليل الرسوم والمخططات البيانية بكل سهولة ودون صعوبات[51]

تم تطوير تطبيق (weka) في مخابر جامعة (Waikato) في نيوزيلندا بالاعتماد على لغة البرمجة جافا وكانت النسخة الأولى في عام 1996، ويعتبر أحد أفضل برمجيات التنقيب في

[50] (الضاهر، 2014)
[51] (www.cs.waikato.ac.nz, n.d.)

البيانات حيث يعمل على كافة أنظمة التشغيل (Windows, Linux, Macintosh) كما يمكن تحميل البيانات إلى التطبيق من عدة مصادر تتضمن:

1- الملفات ذات الصيغ المعروفة لهذه التطبيق (arff, xrff, data, csv, bsi).

2- مختلف قواعد البيانات: (SqlServer, MySQL, Oracle).

3- مسارات ونطاقات عناوين الحواسب ضمن الشبكات الحاسوبية (URLs).

❖ لتدريب الجهاز على تحليل البيانات، تحتاج إلى عدة اعتبارات حول البيانات:

- يجب أن تكون البيانات نظيفة.
- يجب ألا تحتوي على قيم فارغة.

إلى جانب ذلك، لن تكون جميع الأعمدة في جدول البيانات مفيدة لنوع التحليلات التي تحاول تحقيقها يجب إزالة أعمدة البيانات أو "الميزات" غير ذات الصلة كما هو موضح في مصطلحات تعلُّم الآلة، قبل إدخال البيانات في خوارزمية تعلُم الآلة.

باختصار، تحتاج بياناتك الضخمة إلى الكثير من المعالجة المسبقة قبل استخدامها لتعلم الآلة بمجرد أن تصبح البيانات جاهزة، يمكنك تطبيق العديد من خوارزميات التعلم الآلي مثل التصنيف والانحدار والتجميع وما إلى ذلك لحل المشكلة في نهايتك.

يعتمد نوع الخوارزميات التي تطبقها بشكل كبير على معرفة المجال الخاص بك حتى داخل نفس النوع، على سبيل المثال التصنيف، هناك العديد من الخوارزميات المتاحة قد ترغب في اختبار الخوارزميات المختلفة ضمن نفس الفصل لإنشاء نموذج فعال للتعلم الآلي أثناء القيام بذلك، تفضل تصور البيانات المعالجة وبالتالي تحتاج أيضًا إلى أدوات تصور.

يضم هذا التطبيق مجموعة كبيرة من أدوات التنقيب في المعطيات والتعلم الآلي مجانية المصدر وتتضمن الواجهة الرسومية العمليات الرئيسية التالية:

- تحضير البيانات (Pre-processing Data)
- التصنيف (Classification)
- العنقدة (Clustering)
- قواعد الارتباط (Association Rule)

2-1-5- شرح محتوى البيانات المدروسة

تم الحصول على البيانات من البنك البرتغالي "Banco Português de Investimento" نتيجة حملة تسويقية مباشرة عبر (الهاتف) ودامت فترة الحملة لسنة كاملة وتم من خلالها حفظ بيانات العملاء وبيانات الحملة بشكل عام، وتنقسم البيانات إلى (أربع) أقسام وتترتب كالتالي:

1. قسم خاص ببيانات العميل.
2. قسم متعلق بآخر جهة اتصال للحملة الحالية.
3. قسم متعلق بالحملة التسويقية الحالية والحملات السابقة.
4. ويتم إثراء البيانات بل إضافة إلى خمس سمات اجتماعية واقتصادية جديدة (مؤشرات وطنية واسعة من بلد يبلغ عدد سكانه 10 ملايين نسمة تقريبًا)، نشرها مصرف (Banco de Portugal) ومتاح للجمهور بشكل كامل من على الموقع وهي (معدل تغير العمالة، مؤشر أسعار المستهلكين، مؤشر ثقة المستهلك، عدد العاملين، سعر يوريبور 3 أشهر)

بحيث أن إضافة السمات الاجتماعية والاقتصادية الخمس الجديدة (المتاحة هنا) تؤدي إلى تحسن كبير في توقع النجاح من خلال ربط السمات اجتماعية واقتصادية للمجتمع البرتغالي الخارجة من قبل البنك المركزي لتوقع إيداعات العملاء في البنوك الخاصة، بالإضافة سمة الهدف المتغير(Y) الرئيسية التي تعبر عن سمة الإخراج التي تميز بين العملاء من اشترك بوديعة ومن لم يشترك، وبل تالي يكون متغير الإخراج الهدف المطلوب (هل سيشترك في وديعة قام العميل "نعم/لا").

قد شملت البيانات (23) سمة (Attribute)، مقسمة على (أربع) اقسام بحيث تضمنت سمات خاصة بل عميل وشملت (9) سمات، وسمات خاصة بآخر جهة اتصال للحملة الحالية وقد شملت (4) سمات، وسمات خاصة بالحملة التسويقية الحالية والحملات السابقة وقد شملت (4) سمات، وسمات اجتماعية واقتصادية وشملت (5) سمات، بل إضافة الى سمة وعي سمة واحدة (1) تعبر عن إخراج الهدف المطلوب (هل سيشترك العميل في وديعة قام العميل "نعم/لا").

وقد شملت الحملة (41,188) عميل موزعين على أشهر السنة، أي (41,188) سطر (record)، وسيتم شرح مضمون كل سمة (Attribute) على حدة من خلال الجدول التالي:

- السمات الخاصة بالعميل

نوع السمة	السمة(Attribute)	الوصف	المحتوى
السمات الخاصة بالعميل	(ID)	رقم العميل داخل البنك	(numeric)
	(age)	عمر العميل	(numeric) From 17 to 98
	(Job)	عمل العميل	(categorical) • admin • blue-collar • entrepreneur • housemaid • management • retired • self-employed • services • student • technician
	(marital)	الحالة الزوجية للعميل	(categorical) • Divorced • Married • Single • unknown
	(education)	تعلم العميل	(categorical) • basic.4y • basic.6y • basic.9y • high school

			• illiterate • professional course • university degree • unknown
	(balance)	ميزانية العميل	(numeric) From (− 6847) to 10443
	(default)	التخلف عن السداد هل الائتمان في التخلف عن السداد؟	(categorical) • no • yes • unknown
	(Housing)	قرض سكني هل لديه قرض إسكان؟	(categorical) • no • yes • unknown
	(loan)	قرض مالي هل لديه قرض شخصي؟	(categorical) • no • yes • unknown

- السمات الخاصة بآخر جهة اتصال للحملة الحالية

نوع السمة	السمة(Attribute)	الوصف	المحتوى
السمات الخاصة بآخر جهة اتصال للحملة الحالية	(contact)	نوع الاتصال	(categorical) • cellular • telephone
	(month)	آخر شهر اتصال من السنة	(categorical) (jan, feb, mar,..nov,dec)
	(day of week)	آخر يوم اتصال في الأسبوع	(categorical) (mon, tue, wed, thu, fri)
	(duration)	آخر مدة اتصال بالثواني	(numeric) From 0 to 4918 sec

➢ **<u>ملاحظة هامة:</u>**

سمة المدة (duration):

تؤثر هذه السمة بشكل كبير على هدف الإخراج كالتالي:

" على سبيل المثال، إذا كانت مدة المكالمة = (0) فسوف يكون حتما هدف الإخراج = (no) " ومع ذلك، فإن المدة غير معروفة قبل إجراء المكالمة أي لا يمكن للموظف المتصل معرفة المدة التي سيتكلم بها قبل إنهاء المكالمة، أيضا بعد نهاية المكالمة يكون هدف الإخراج معروف بوضوح، فإن ادخال سمة المدة على النماذج التنبؤية يمكن أن تضعف دقة المخرجات بشكل كبير فإن كانت المسألة مسألة تصنيف فسوف تنقص دقة المصنف بشكل كبير إذا ادخلت هذه السمة، وبالتالي لا ينبغي إدراج هذا السمة إلا لأغراض قياس الأداء وتصوير البيانات ويجب التخلص منه إذا كان القصد هو الحصول على نموذج تنبؤي واقعي.

- السمات الخاصة بالحملة التسويقية الحالية والحملات السابقة

نوع السمة	السمة(Attribute)	الوصف	المحتوى
السمات الخاصة بالحملة التسويقية الحالية والحملات السابقة	(campaign)	عدد جهات الاتصال التي تم إجراؤها خلال هذه الحملة وبالنسبة لهذا العميل(يشمل آخر اتصال)	(numeric) From 1 to 56
	(pdays)	عدد الأيام التي مرت بعد آخر اتصال للعميل من حملة سابقة (999): يعني أنه لم يتم الاتصال بالعميل سابقًا	(numeric) From 999 to 27
	(previous)	عدد جهات الاتصال التي تم إجراؤها قبل هذه الحملة وبالنسبة لهذا العميل	(numeric) From 0 to 7
	(poutcome)	نتيجة الحملة التسويقية السابقة	(categorical) • Failure • Nonexistent • success

- السمات الاجتماعية الاقتصادية

المحتوى	الوصف	السمة(Attribute)	نوع السمة
(numeric) المؤشر ربع السنوي	معدل تغير العمالة	Emp.var.rate	السمات الاجتماعية والاقتصادية
(numeric) المؤشر الشهري	مؤشر أسعار المستهلك	Cons.price.idx	
(numeric) المؤشر الشهري	مؤشر ثقة المستهلك	Cons.conf.idx	
(numeric) المؤشر اليومي	سعر يوريبور (3) أشهر	Euribor(3)month	
(numeric) المؤشر ربع السنوي	عدد الموظفين	nr.employed	

➢ **<u>شرح المؤشرات الاجتماعية والاقتصادية:</u>**

معدل تغير العمالة: (Emp.var.rate)

يحدد مستوى التغير في مستوى الناس الموظفين خلال الربع سنوي الماضي من العام. يعتبر خلق الوظائف مؤشر هام للتعافي الاقتصادي حيث يؤلف الاستهلاك، والذي يرتبط بشكل كبير بأوضاع الموارد البشرية، جزء كبير من الإنتاج المحلي الإجمالي للاتجاه التدريجي بتأثير إيجابي على عملة البلاد.

مؤشر أسعار المستهلك: (Cons.price.idx)

مؤشر أسعار المستهلك هو رقم استدلالي أو مؤشر إحصائي يقيس التغيرات التي تحصل في المستوى العام للأسعار، انطلاقا من تتبع سلة تشمل جميع السلع والخدمات المستهلكة داخل بلد معين، ويُفترض في تركيبة هذه السلة أن تعكس بنية الإنفاق الاستهلاكي للأسر في هذا البلد

مؤشر ثقة المستهلك: (Cons.conf.idx)

مؤشر ثقة المستهلك من المؤشرات الاقتصادية الهامة، ويهدف هذا المؤشر الى قياس رأي المستهلك حول الوضع الاقتصادي له ولأسرته الحالي وتوقعاته المستقبلية نحو أداء الاقتصاد بشكل عام، على المدى القصير والمتوسط بالإضافة إلى قدرته ورغبته في شراء السلع مثل الأجهزة والسيارات والأثاث والمنازل.

سعر يوريبور (3) أشهر: (Euribor 3)

ليوريبور (أو سعر الفائدة على معاملات اليورو ما بين المصارف) هو السعر الذي يقبل بموجبه مصرف رئيسيّ إقراض الأموال باليورو لمصرف رئيسيّ آخر. يتمّ احتساب اليوروبور يوميّاً لرصد الودائع ما بين المصارف التي تبلغ فترة استحقاقها أسبوعاً واحداً وأسبوعاً واحدة إلى 12 شهراً كمعدّل لأسعار العرض اليوميّة لفريق ممثلي المصارف الرئيسيّة، ويتمّ تقريبه إلى ثلاث خانات عشريّة.

- متغير الإخراج

(هل سيشترك في وديعة قام العميل "نعم/لا")

المحتوى	الوصف	السمة(Attribute)	نوع السمة
(categorical) • Yes • no	(هل سيشترك في وديعة قام العميل "نعم/لا")	(Categorical/ Binary)	هدف الإخراج

1-6- المبحث الثاني

1-1-6- تصوير البيانات (Data Visualize)

يشير تصور البيانات إلى التقنيات المستخدمة لإيصال الأفكار من البيانات من خلال التمثيل البصري، وهدفها الرئيسي هو تقطيع مجموعات البيانات الكبيرة إلى رسومات مرئية للسماح بفهم سهل للعلاقات المعقدة داخل البيانات، وغالبًا ما يتم استخدامه بالتبادل مع مصطلحات مثل رسومات المعلومات والرسومات الإحصائية وتصور المعلومات، وهي إطار للتعامل مع مهام علوم البيانات بعد جمع البيانات ومعالجتها ونمذجتها، يجب تصور العلاقات حتى يمكن التوصل إلى نتيجة.

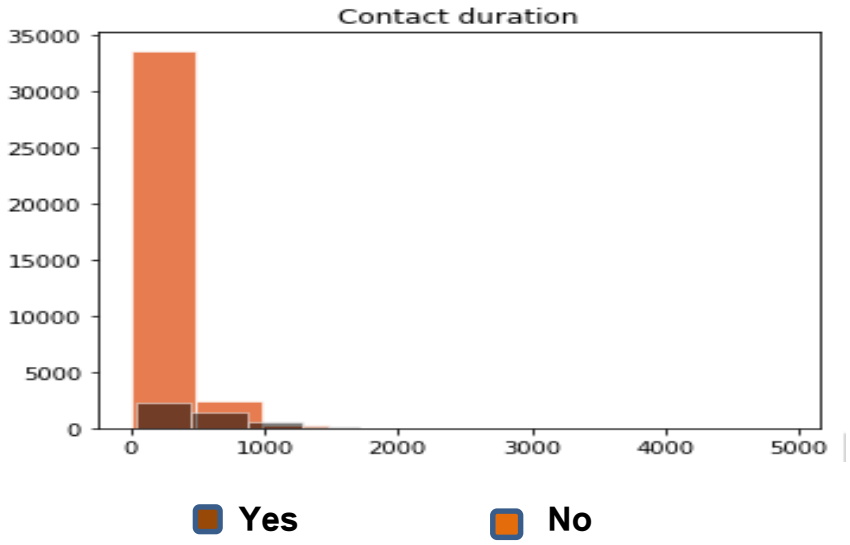
وهو أيضًا أحد مكونات النظام الأوسع لهندسة عرض البيانات (DPA)، والذي يسعى إلى تحديد البيانات وتحديد موقعها ومعالجتها وتنسيقها وتقديمها بأكثر الطرق فعالية.

- لدينا نوعان من المتغيرات في مجموعة البيانات لدينا قسم رقمي (numeric)، وقسم فئوي (categorical).

❖ **المتغيرات الرقمية**

(age) و(Duration) و (campaign) و (pdays) و (previous) و (emp.var.rate) و (cons.price.idx) و (cons.conf.idx) و (euribor3m) و (nr.employed) وسوف يتم تصوير أهمها:

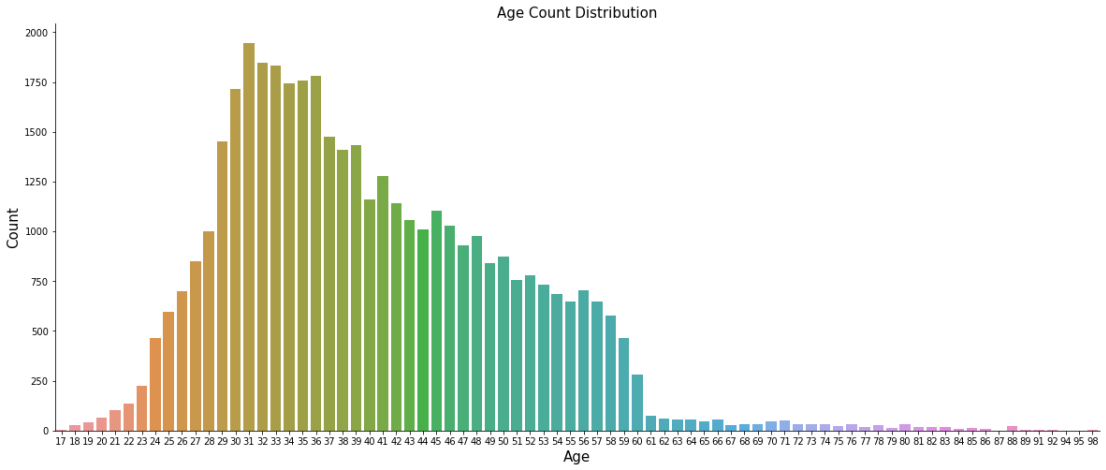
- **المدة: (Duration)**



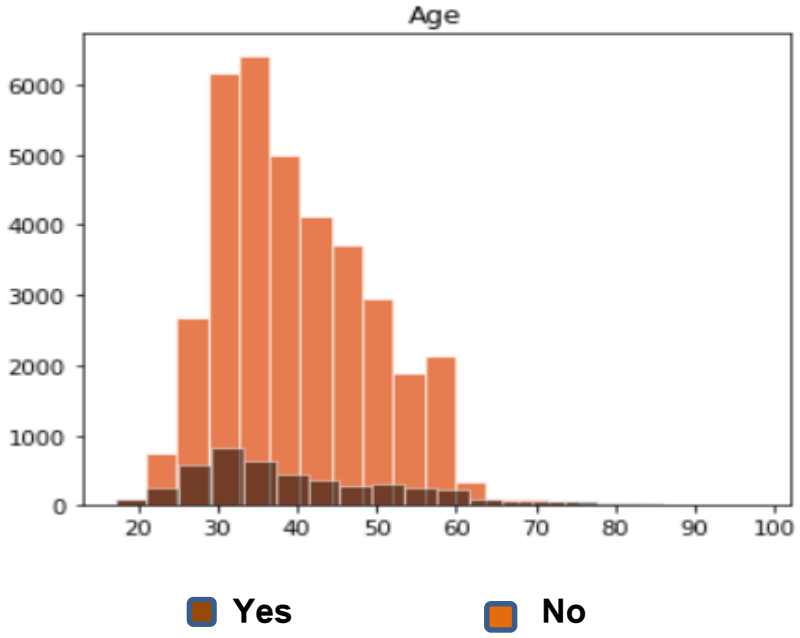
وهنا يظهر لنا أن العملاء الذين يتصل بهم 20 دقيقة (60/1200) لا يتفاعلون مع الحملة التسويقية بتاتا، وإن العملاء الذين يتصل بهم 8 دقائق (60/500) هم اغلب العملاء الذين لا يودعون يليهم العملاء الذين يتصل بهم 16 دقيقة (60/100) لكن بنسبة أقل بكثير، ومن ثم العملاء الذين يتصل بهم 25 دقيقة هم فئة قليلة جدا ولم يودعوا نسبيا.

مقارنتا بل عملاء الذين أودعوا فهم متوزعين بين ال 0 إلى 1300 ثانية بشكل يشابه لكن أقل بقليل من الذين لم يودعوا، لكن هناك مكان يتقارب منه العملاء الذين أودعوا والذين لم يودعوا وهو من ال 600 إلى 900 ثانية أي 9 إلى 14 دقيقة فيجب تكثيف الحملة التسويقية للوصول لهذه المدة من المكالمات بحيث تشمل جميع العملاء وليس القلة منهم لأن حسب الرسم البياني تبين أن الاتصال لمدة تتراوح ما بين ال 10 إلى 15 دقيقة هي انسب مدة للمكالمة لكن طبعن مثلما ذكرنا سابقا لا يجب أن نحسب المكالمات التي مدتها = (0) لأن هدف الإخراج معروف مسبقا وهو (no)، لكن يجب التقيد بهذا العدد من الدقائق للحملات القادمة قدر المستطاع وتجنب المكالمات التي تعلو مدتها عن ال 14 دقيقة ولا تكون أقل من 9 وعلى ذلك يمكن تخفيض تكلفة الحملات من خلال ضبط مدة المكالمات التي تطول فوق المدة المطلوبة لنجاح إيداع العملاء، وبما أن الحملة تطبق من خلال المكالمات فيجب أخذ بعين الاعتبار المدة الصحيحة فهي عامل الرئيسي لنجاح نتائج الحملات ويجب ربطها بباقي البيانات لتعطي تأكيدا أكبر لنجاح الإيداعات.

- **العمر (age)**

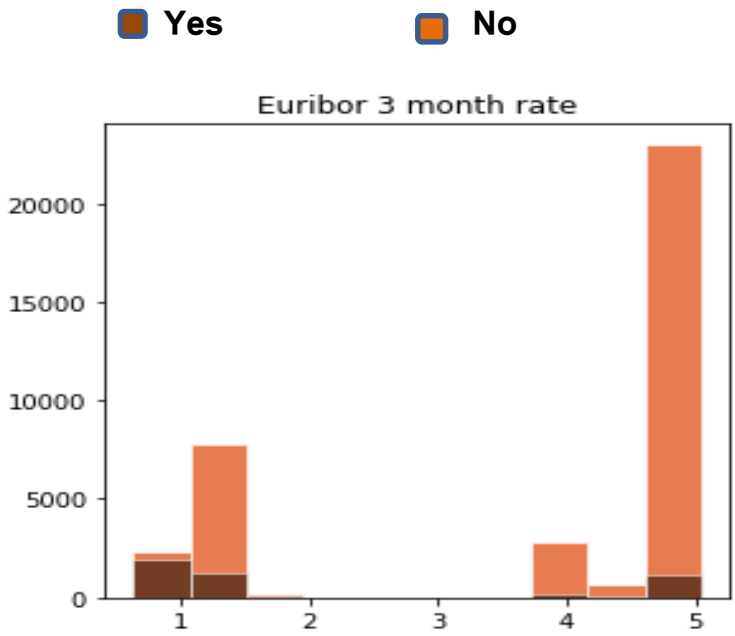


ويظهر الرسم البياني انتشار الاعمار بين ال (17) سنة إلى (98) سنة، وأن اكثر فئة عمرية من العملاء هي بين (30) إلى (45) سنة، وإن الاعمار من ال (17) إلى (20) ومن ال (60) فما فوق عم أصغر فئة.



يظهر الرسم فئات الاعمار للعملاء الذين اشترك بودائع ومن لم يشترك ويوضح الرسم البياني أن مجمل الذين اشتركوا بوديعة استثمارية هم من جميع الفئات لكن تركزت عند العملاء التي أعمارهم بين ال 30 و40 فيجب استهدافهم أكثر في الحملات القادمة من خلال استهداف هذه الفئة بمدة مكالمة معينة قد ذكرت سابقا لنجاح الحملة ورفع نسبة الايداعات لباقي الشرائح، وبالنسبة للعملاء الذين عمرهم ما فوق (60) هم قلة لكن أغلبهم قد اشترك بوديعة.

- سعر يوريبور (3) أشهر: (Euribor (3) month

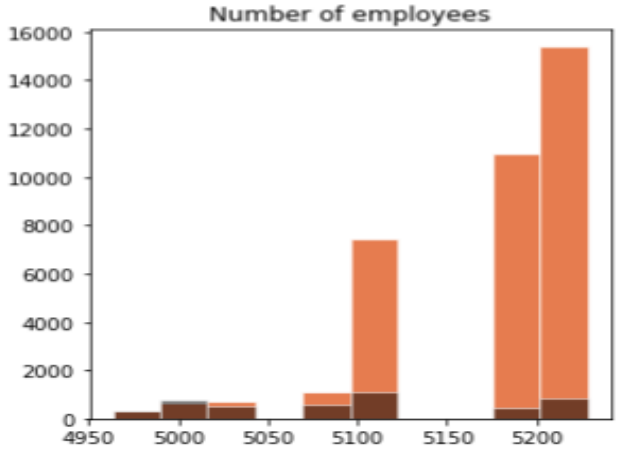


يظهر من خلال الرسم البياني أن فئة أكثر العملاء الذين لم يودعوا كانوا عندما كان المؤشر قريب من الرقم (4.6) إلى (5.1)، وأن بين رقم (2) إلى (3.6) لم يتفاعلوا مع الحملة بتاتا، وإن غالب العملاء الذين أودعوا هم ما بين (0.5) إلى (1.6)، وكما ذكرنا أن ليوريبور أو سعر الفائدة على معاملات اليورو ما بين المصارف هو السعر الذي يقبل بموجبه مصرف رئيسيّ إقراض الأموال باليورو لمصرف رئيسيّ آخر، أي أنه يتوافق مع المنطق الاقتصادي فكل ما كان سعر الفائدة على القروض أقل فسوف تتحسن حركة الإيداعات بشكل عام في كل البنوك، ومن هنا فيجب على الحملة تكثيف المكالمات عندا يكون المؤشر ما بين (0.5) إلى (1.6) وبما أن هذا المؤشر يومي فيجب تكثيف المكالمات عندا تكون الأيام عند المؤشرات السابقة لنجاح الحملة التسويقية، وإن هذا المؤشر مهم جدا بقرار إيداعات العملاء فمن المهم اخذ بعين الاعتبار تلك الأيام التي أظهر لنا الرسم البياني كثافة اشتراكات العملاء بالودائع.

- **مؤشر ثقة المستهلك: (cons.conf.idx)**

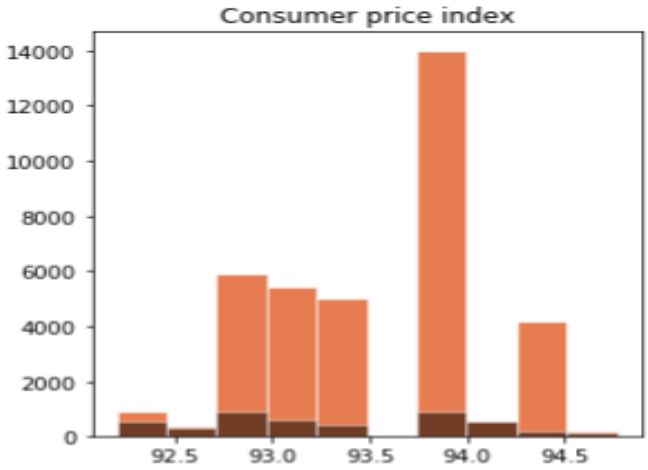
بما أن هذا المؤشر (شهري) فيظهر الرسم البياني توزع معدلات المؤشر على أشهر السنة موضحا ارتفاع الحملة في الأشهر بالترتيب (6) (9) (4) على الرغم من عدم إيداع العملاء مقارنة بباقي الأشهر، وأن في الأشهر (3) (7) (8) (10) قد كانت الحملة منخفضة لكن ترافق معها حركة إيداعات شبه كبيرة، وبما أن ثقة هذا المؤشر يتمحور حول رأي العميل بالوضع الاقتصادي وقدرة شرائه للمنتجات فيظهر ارتفاع احتمال إيداعه بالأشهر السابقة ويجب تكثيف الحملة بهم، وتخفيضها بتلك الأشهر التي لا تأتي بإيداعات، وهذا يتوافق مع المنطق الاقتصادي بأن قدرة الفرد العميل على الشراء ورأيه وأمانه يأثر على نسبة إيداعه

- **عدد العاملين:**



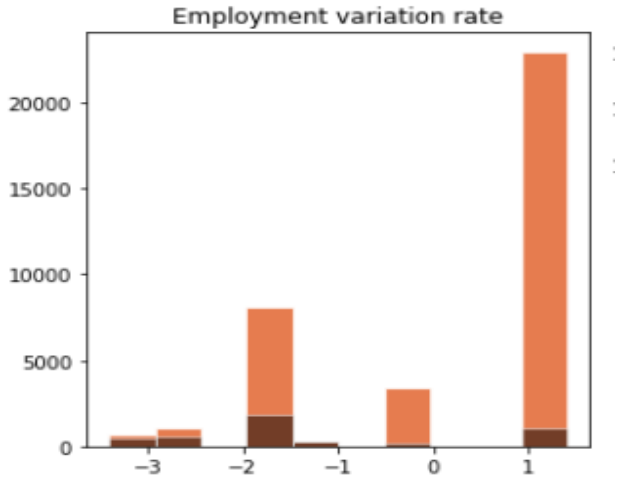
يظهر الرسم البياني أن أفضل عدد للموظفين ارتباطا مع نجاح الحملة التسويقية هوة (5200)

- **مؤشر أسعار المستهلكين: (cons.price.idx)**



يظهر الرسم البياني التوزع لمؤشر أسعار المستهلكين (مؤشر شهري) أن الحملة كثفت بشكل كبير عندما كان المؤشر يقارب (93.8) إلى (94) ويوجد بهذا الشهر نتائج سالبة للحملة مقارنة بالعملاء الذين أودعوا، فيجب تكثيف الحملة على باقي الأشهر وخاصة الشهر الرابع والخامس والحادي عشر لأنه يملك عمليات إيداع فقط وهذا ما تحتاجه الحملة التسويقية، وهذا يتفق مع بأن ارتفاع الأسعار يؤثر بشكل كبير على إيداعات الأفراد والعكس صحيح.

- مؤشر تغير العمالة: (mp.var.r)

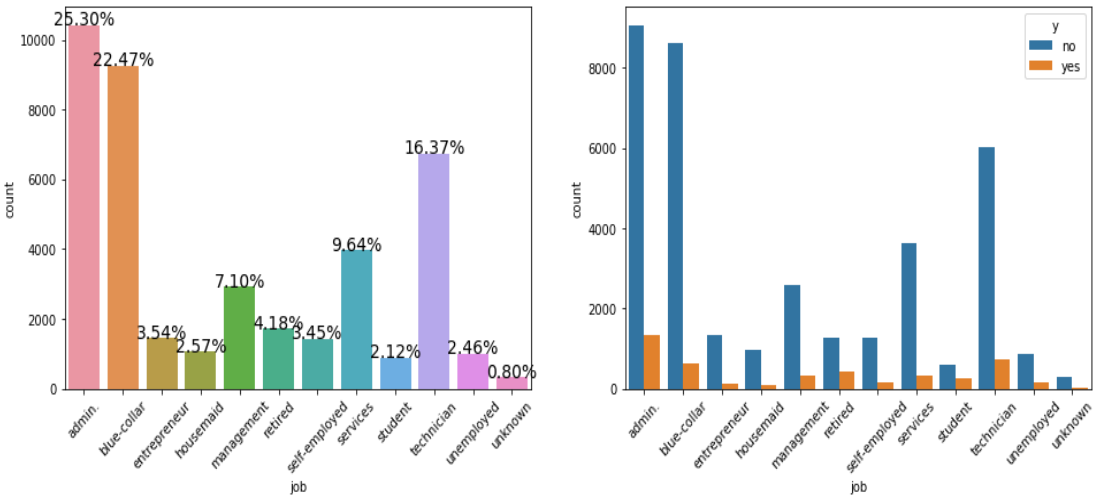


يظهر الرسم البياني لمؤشر تغير العمالة (مؤشر فصلي) تكثيف الحملة التسويقية في الفصل الثاني من السنة وفي الشهر الأخير تحديدا مقارنتا بل فصل الأول من السنة، فيجب على الحملة تكثيف المكالمات في الفصل الأول من السنة، وأيضا يتوافق مع المنطق الاقتصادي بحيث كثرة العمالة تؤدي إلى حركة إيداعات أكثر والعكس صحيح

❖ **المتغيرات الفئوية: (categorical)**

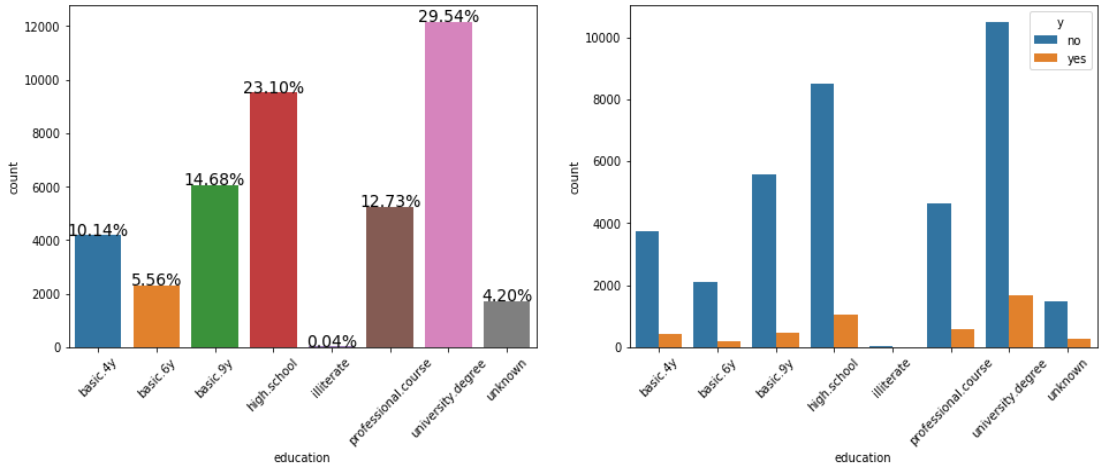
(Job) و (Month) و(Marital) و (Default) و (Housing) و(Loan) و(Contact) و (Poutcome) و (Day of week) و(y) و (E)

- **العمل: (Job)**



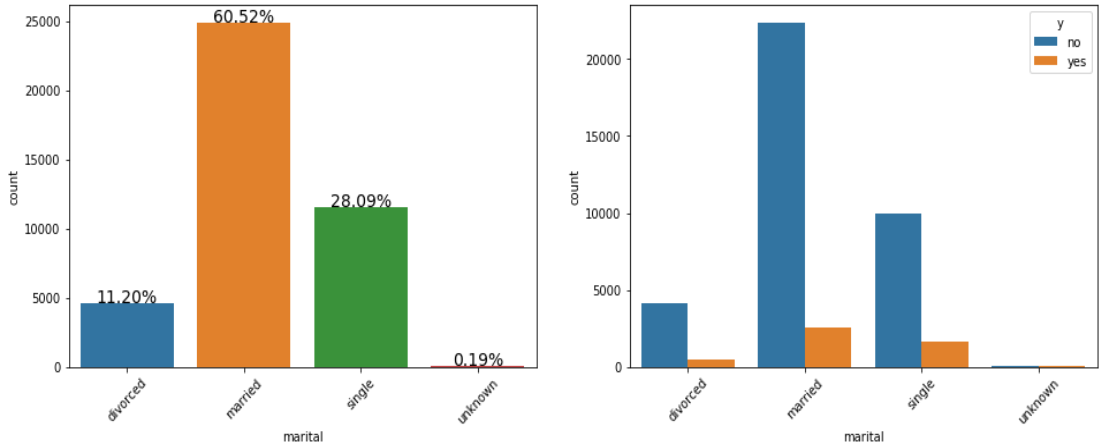
يظهر من خلال الرسم البياني أن أكبر فئة من العملاء الذين شملتها الحملة التسويقية هم بالترتيب (Admin) و (Blue-Coler) و (Technician)، وأن أصغر الفئات هم (Student) و(Unknown)، وأيضا هم أكثر الفئات التي اشتركوا في الودائع فيجب تكثيف الحملة بحيث عليهم.

- **التعليم: (Education)**



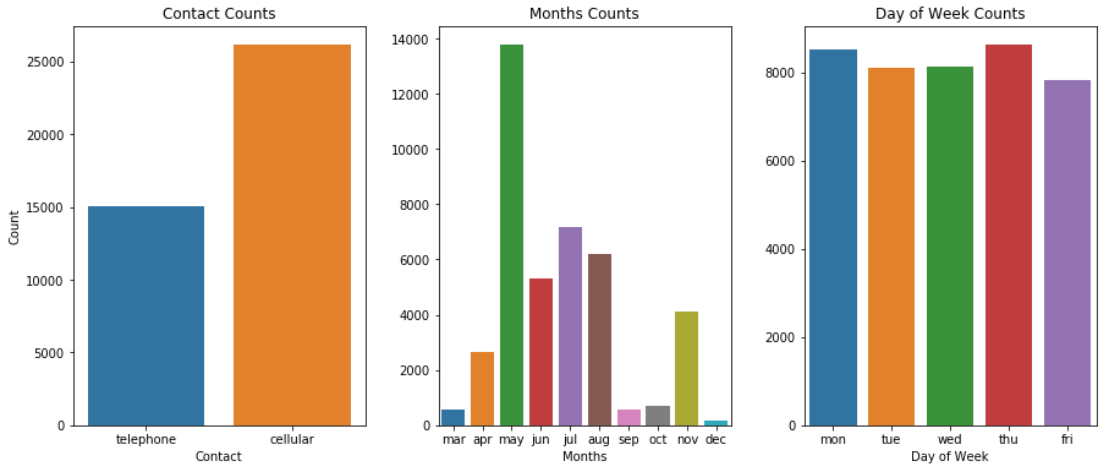
يظهر من خلال الرسم البياني أن أكبر فئة من العملاء الذين شملتهم الحملة بالترتيب هم (University degree) و (High school)، الفئات الصغيرة هم (Basic 9y) و (Unknown)، وأن الفئتين (University degree) و(High school) هم أكثر فئتين اشتركوا بالودائع البنكية.

- **الحالة الزوجية: (Marital)**



يظهر من خلال الرسم الباني أن أكبر فئة من العملاء الذين شملتهم الحملة التسويقية كانوا بأكبر نسبة من (Married)، وبعدها بنسبة اقل هم (Single)، وعدها (Divorced)، وبنسبة شبه معدومة هم (Unkown)، وأيضا الفئة (Married) هي أمثر فئة تفاعلت مع الحملة من حيث الإيداع فيجب تكثيف الحملات المستقبلية عليهم.

- **الشهر (Month) / نوع الاتصال (Contact) / آخر يوم اتصال في الأسبوع (Day of Week)**



- **الشهر (Month):**

يظهر الرسم البياني للأشهر بأن الشهر (May) هو أكثر شهر تم التحدث فيه مع العملاء، يليه شهر (Jul)، ثم شهر (Aug)، وأكثر أشهر كانت الحملة بها ضعيفة هم (Mar) و (Sep) و(Oct)، وفي الشهر الأخير كانت شبه متوقفة.

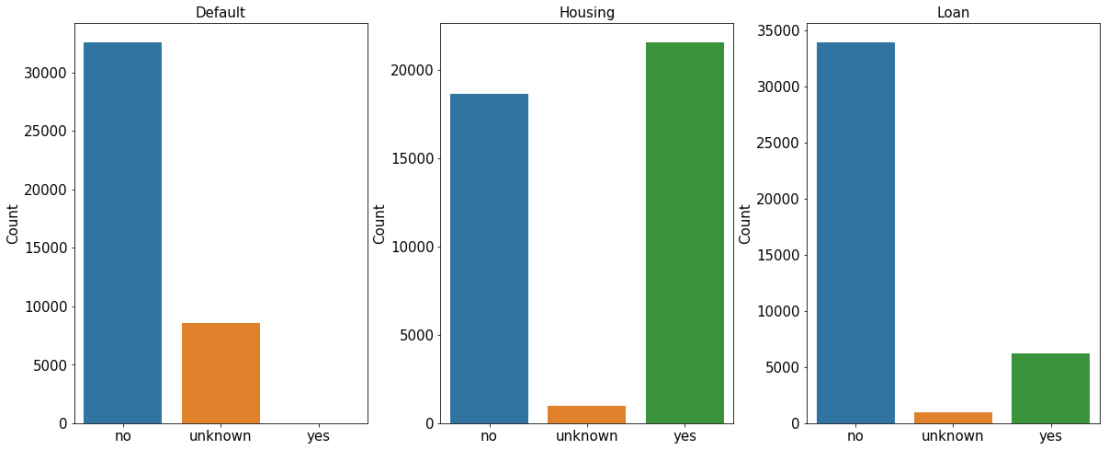
- **نوع الاتصال (Contact):**

يظهر الرسم البياني أن نوع الاتصال مائل أكثر إلى (Cellular) مقارنتا ب (Telephone) أي أن الحملة اتجهت حول الاتصال عبر ال (Cellular).

- **آخر يوم اتصال في الأسبوع (Day of Week)**

يظهر الرسم البياني أن الحملة التسويقية بشكل عام تكلمت بجميع الأيام بشكل شبه متساوي أي ليس هناك أيام تتفوق على أيام بالنسبة لعدد المكالمات مع العملاء.

- **القرض (Loan) / القرض السكني (Housing) / التخلف عن السداد (Default)**



- **القرض (Loan):**

يظهر الرسم البياني أن العملاء الذين تم الاتصال بهم جميعا كانوا بشكل كبير جدا لا يملكون القروض في هذا البنك أو بي البنوك الأخرى (no)، وبنسبة صغيرة منهم كانوا يملكون (Yes)، وفئة ضئيلة جدا هم غير معروف إذا كانوا يملكون أم لا(Unknown).

- **القرض السكني (Housing):**

يظهر الرسم البياني أن العملاء الذين قد شملتهم الحملة التسويقية نسبة كبيرة منهم يملكون قروض سكنية (Yes)، وأيضا نسبة كبيرة لا يملكون (No) لكن الفئة التي تملك هي الغالبة، ونسبة ضئيلة هم غير معروفين (Unknown).

- **التخلف عن السداد (Default):**

يظهر الرسم البياني أن العملاء الذين قد شملتهم الحملة التسويقية نسبة كبيرة جدا غير متخلفين عن السداد (No)، ونسبة صغيرة غير معروفة (Unknown)، ونسبة ضئيلة جدا هم متخلفين (Yes).

2-1-6- اختيار التقنيات المستخدمة

إن مشكلة البحث تتمحور حول التنبؤ بنتيجة الحملات التسويقية المباشرة لتوقع إيداعات العملاء في البنك، فعندما نتكلم على التوقع أو التنبؤ غالبا ما تكون التقنيات المستخدمة تحت تقنيات التصنيف والذي يدلي تأكيدا لذلك هي نوعية البيانات المستخدمة والتي تضم سمة (Attributes) إخراج من نوع (نعم / لا) وفقا لإيداعات العملاء في البنك وهذا يؤكد اختيارنا لتقنية التصنيف وما تضم من خوارزميات ومصنفات عدة، فسوف يتم استخدام (5) مصنفات تابعة لتقنية التصنيف وهم (" الغابات العشوائية (Random Forest)، شجرة القرار (Decision Tree J48)، الانحدار اللوجيستي (logistic regression)، الجار الأقرب (KNN)، وبايز الساذج (naive bayes) ")، وسيتم تطبيق المصنفات على البيانات المستخدمة من خلال أداة ويكا (WEKA) التي تكلمنا عنها في المبحث الأول، ولكن قبل تطبيق المصنفات يجب تحضير هذه البيانات بالشكل الصحيح مثلما ذكرنا في الفصل الثالث بخطوات تحضير البيانات من خلال عدة خطوات تحتاجها التقنية أن تحضر بها البيانات المدخلة كي تتم عملية التصنيف بالشكل الصحيح وإعطاء دقة للمصنف المخرج، وبالفقرة القادمة سيتم تحضير البيانات المدروسة لتطبيق تقنية التصنيف عليها.

3-1-6- بناء المصنفات

1-3-1-6- تحضير البيانات

تعتبر المعالجة المسبقة للبيانات خطوة مهمة في أي مشروع بهدف التنقيب في هذه البيانات، فبيانات العالم الحقيقي بشكل عام غير مكتملة وتحوي أخطاء، من أجل إنتاج نموذج جيد لحل هذه المشكلة يجب معالجة هذه القضايا واخذها بعين الاعتبار سنستعرض فيما يلي عدة مسائل وتقنيا مستخدمة لتجهيز البيانات:

1) تنظيف البيانات:

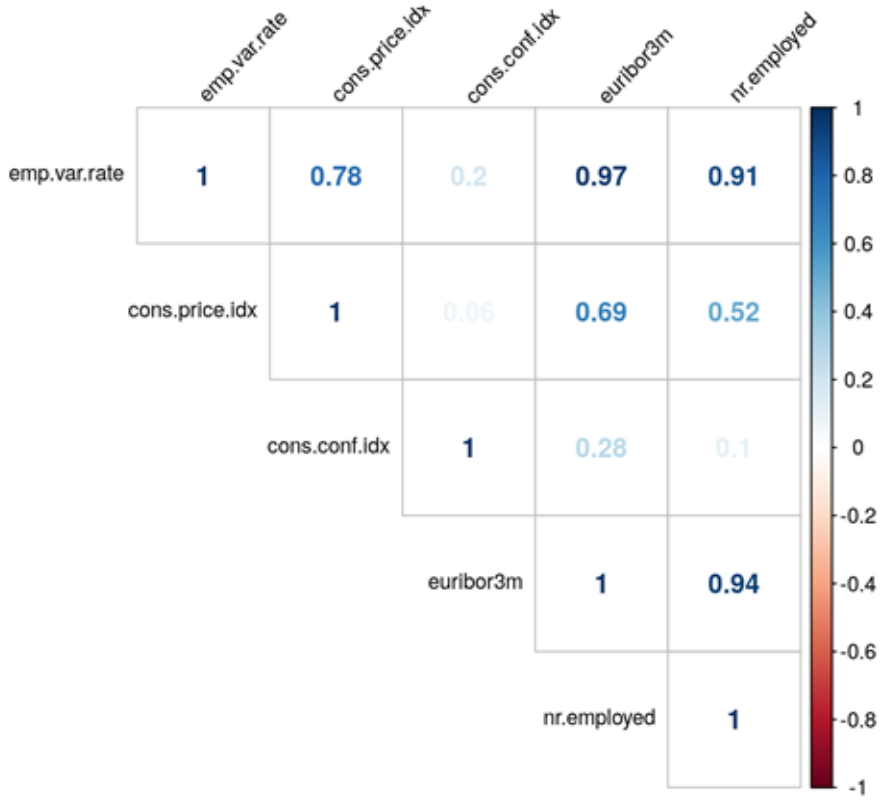
في هذه المرحلة يتم تنظيف البيانات من خلال حذف السمات التي لا تدلي بمعلومات، أو تنظيف الخلايا الفارغة إما عن طريق الحذف أو عن طريق تعبئتها بقيم تتمثل بمتوسط باقي الخلايا في نفس السمة، لكن البيانات المدروسة لا تحتوي على خلايا فارغة، لكن يمكن القيام بحذف سمة واحدة تابعة لبيانات العميل وهي (ID) التي تضم رقم العميل التابع لدى البنك ويفيد حذفها لأنها يمكن أن تؤثر على دقة المصنف بسبب محتواها التي لا قيمة له في عملية التنبؤ، ومنه سوف يصبح عدد السمات هو (22) سمة.

2) تحليل مدى الارتباط: (correlation analysis)

قد تحتوي البيانات في بعض الأحيان على مجموعات قليلة من الصفات التي لا توفر الكثير من المعلومات للمصنف هذا يعني انه لا تشكل هذه السمات إدخالات مهمة في متجه الميزة وإزالة هذه السمات يمكن لهذه الصفات تسريع أداء المصنف ومن خلال (correlation analysis) يمكن إزالة الترابط المخفي بين هذه السمات كالتالي:

1) بالنسبة للسمات (سمات العميل / سمات المتعلقة بأخر مكالمة /سمات المتعلقة بالحملة التسويقية) قد تم عمل تحليل ارتباط خاص بهم وقد أرفقت نسخة منه مع CD البحث ولم يكن هناك ارتباطا فيما بينهم وجميع الارتباطات لا تتعدى (0.6) ولا يجب أن نحذف منهم أي سمة.

2) بالنسبة للسمات الاجتماعية والاقتصادية الخمس المضافة فيجب تحليل ارتباط فيما بينهم لأنه غالبا ما تكون المؤشرات الاقتصادية والاجتماعية مرتبطة فيما بينها ارتباطا وثيقا لأنها تعود إلى معادلات رياضية مشابهة أو تعود إلى أصل واحد مثال (مؤشر ثقة المستهلك ومؤشر أسعار المستهلك)



يظهر المخطط ارتباطات عديدة وسيتم حذف (عدد العاملين) و(مؤشر أسعار المستهلك) و (معدل تغير العمالة) وسيتم الاحتفاظ بالمؤشرين (سعر اليوربيور) و(ثقة المستهلك) لأنهم الأفضل من خلال الرسم البياني والتقسيمات بهم كانت متوازنة مقارنة بالبقية وأن ال (سعر اليوربيور) هو أهم مؤشر من بينهم بالنسبة للبنك لأنه يعطي معلومات ثابتة عند رقم معين من المؤشر ويظهر بالرسم البياني في تصوير البيانات.

" ومنه سوف يصبع عدد السمات جميعا هو (19) "

3) تحويل البيانات: (Data transformation)

- **التعميم: Generalization**

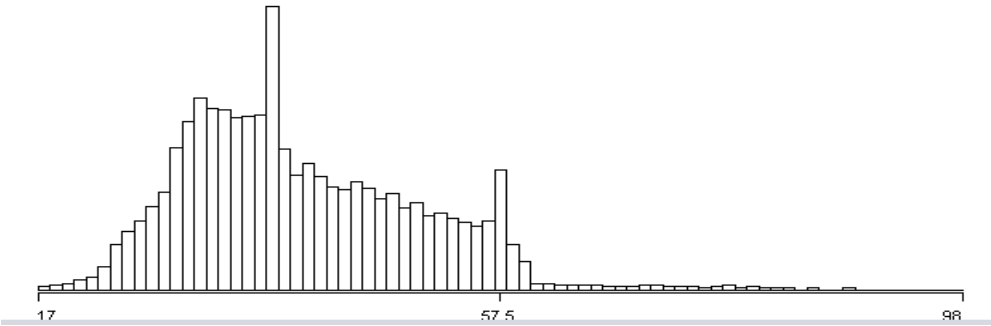
يمكن أيضًا تحويل البيانات من خلال تعميمها على المفهوم الأعلى لهذا الغرض يمكننا استخدام التسلسل الهرمي للمفاهيم. تم تطبيق عملية (تعميم generalization) على البيانات لتوليد سويات أعلى، وباستخدام هنا (المفاهيم الهرمية concepts hierarchies) وهذه العملية مفيدة بالتحديد من أجل الواصفات ذات القيم المستمرة، فعلى سبيل المثال يمكن تعميم القيم الرقمية للواصفة (Balance) على شكل مجال ذي قيم متقطعة مثل: {Low, Medium, High}، أي نحول من السمات الرقمية إلى فئوية

1) العمر (Age):

لقد توزعت اعمار العملاء في الحملة التسويقية من 17 إلى 98 كالتالي:

Statistic	Value
Minimum	17
Maximum	98
Mean	40.024
StdDev	10.421

Class: age (Num) Visualize All



ومن الرسم البياني يظهر لنا تركز الفئات العمرية بطرف والطرف الاخر منخفض جدا أي الفئات العمرية ليست متساوية، ولذلك يجب التقسيم للحصول على تقسيمات متقاربة وتشمل الأماكن

التي ترتكز بها الفئات بتقسيمات صغيرة للحصول على معلومات مفيدة أكثر وتمت التقسيمات كل تالي:

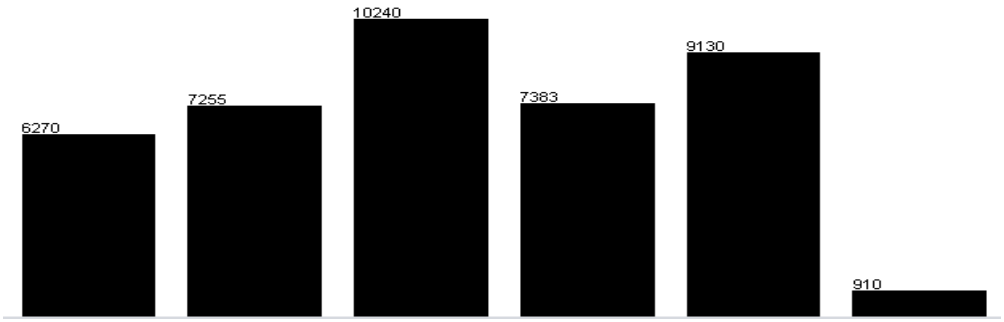
- من 17 إلى 30
- ومن 30 إلى 35
- ومن 35 إلى 40
- ومن 40 إلى 50
- ومن 50 إلى 60
- و60 فما فوق

تنويه: لا يجب اخذ العينة من 50 فما فوق لأن نسب العملاء ما فوق ال 60 يشتركون بودائع بشكل جيد لذلك كانت ال التقسيمات هكذا لما تطلبه البيانات من معلومات.

واصبحت اشكال التوزيعات كالتالي:

No.	Label	Count	Weight
1	from 50 to 60	6270	6270.0
2	from 35 to 40	7255	7255.0
3	from 40 to 50	10240	10240.0
4	from 17 to 30	7383	7383.0
5	from 30 to 35	9130	9130.0
6	more than 60	910	910.0

No class | Visualize All



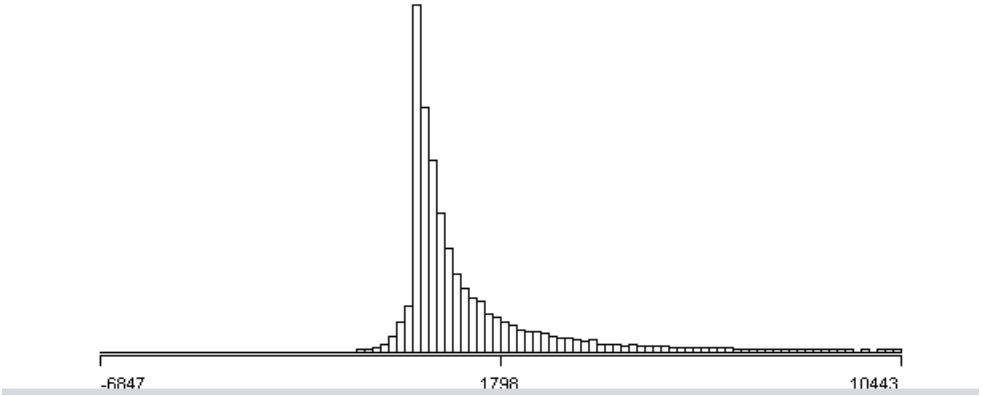
هذا أفضل توزيع توصلت اليه والقما ذكرت الفئة الصغيرة قسمتها هكذا لأنها عملائها على الرغم من قلتهم لكن تجاوبهم مع الحملة التسويقية كان عالي.

2) الميزانية: (balance)

لقد توزعت ميزانية العملاء من (6847-) إلى 10443 على الشكل التالي:

Statistic	Value
Minimum	-6847
Maximum	10443
Mean	1076.498
StdDev	1711.49

Class: balance (Num) Visualize All



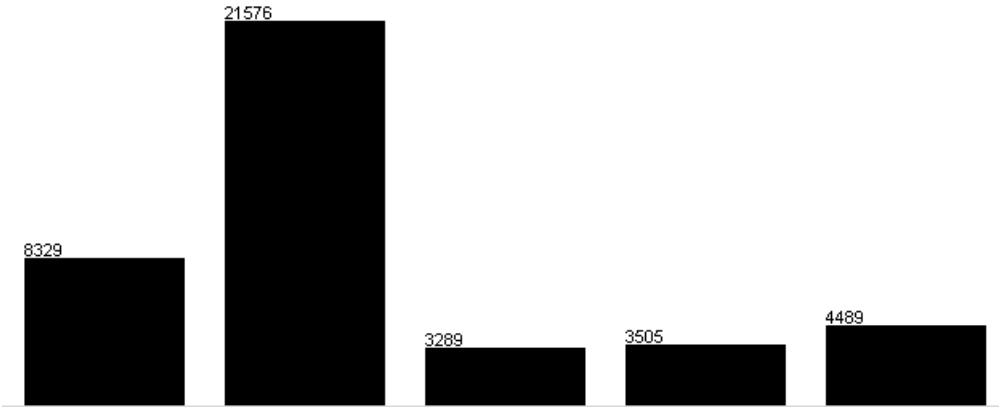
ويظهر في الميزانية الخاصة بالعملاء أن هناك من هو خاسر أي يأتي في لمكان السالب ومنهم من هو متعادل أي (0) ومنهم في الموجب أي يملك ميزانية، وتكتلت توزيعاتهم بين (1000) إلى (2000) لذلك تمت التقسيمات كل تالي:

- من السالب إلى التعادل (LOSS)
- من 1 إلى 1000 (low)
- من 1001 إلى 2000 (medium)
- من 2001 إلى 10443 (high)

وأصبح توزيعهم كالتالي:

No.	Label	Count	Weight
1	medium	8329	8329.0
2	low	21576	21576.0
3	zero	3289	3289.0
4	loss	3505	3505.0
5	high	4489	4489.0

No class | Visualize All

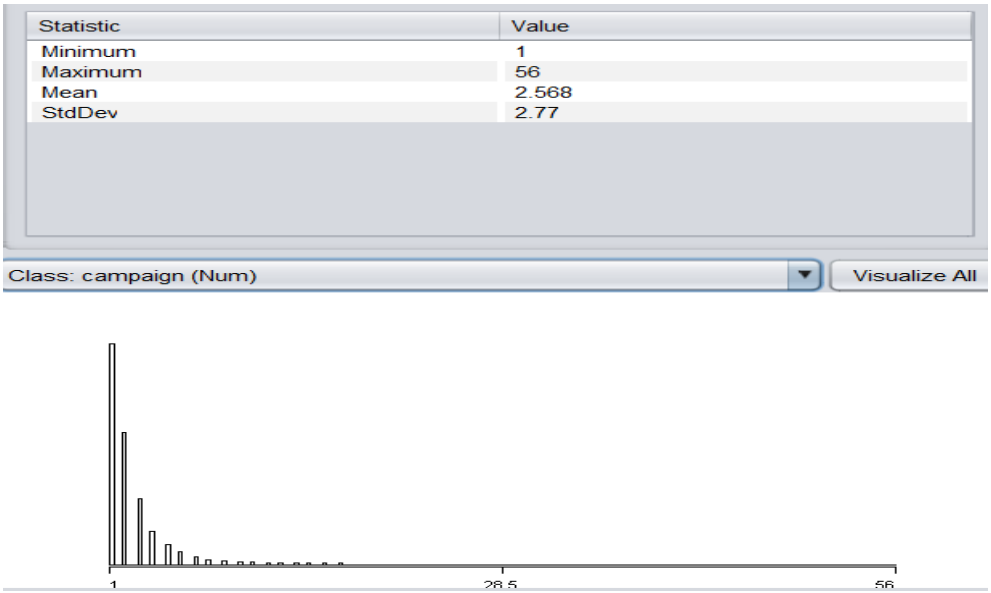


أن أكبر فئة هي فئة (Low) ومن ثم (medium)

تنويه: كما ذكرنا لكل بيانات لها طبيعتها الخاصة وإن مجمل الإيداعات تموضعت في الفئات الصغيرة لذلك، ولا يجب جمع العملاء الذين رصيدهم سالب مع الذين يملكون ميزانية عالية.

3) الحملة (Campaign)

وكان توزيع العملاء الذين قد تم الاتصال بهم بهذه الحملة فقط كالتالي:



ويظهر الرسم البياني أن غالبية العملاء قد تم الاتصال بهم من مرة إلى ثلاث مرات والفئة الغالبة هب الاتصال لمرة واحدة لذلك سيتم توزيع الفئات كالتالي:

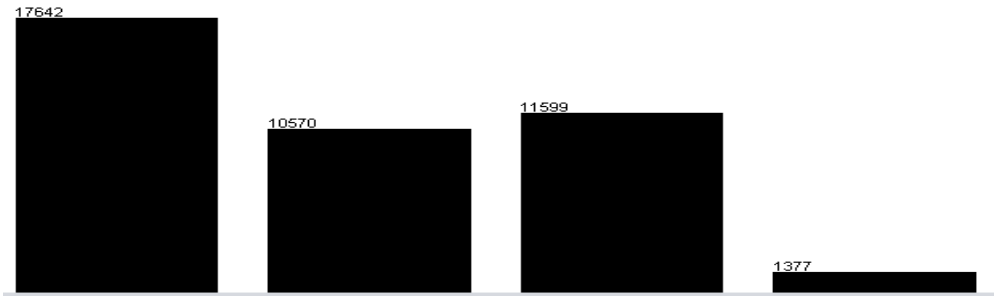
- من 1 إلى 2 اتصال (low)
- من 2 إلى 3 اتصال (medium)
- من 3 إلى 8 اتصال (high)
- من 8 إلى 56 اتصال (very high)

تنويه: حيث لا يمكن قسم المكالمة إلى مكالمة ونصف أو مكالمتين ونصف وهكذا...

وأصبح توزيع الاتصالات على العملاء لهذه الحملة كالتالي:

No.	Label	Count	Weight
1	low	17642	17642.0
2	medium	10570	10570.0
3	high	11599	11599.0
4	very high	1377	1377.0

No class | Visualize All

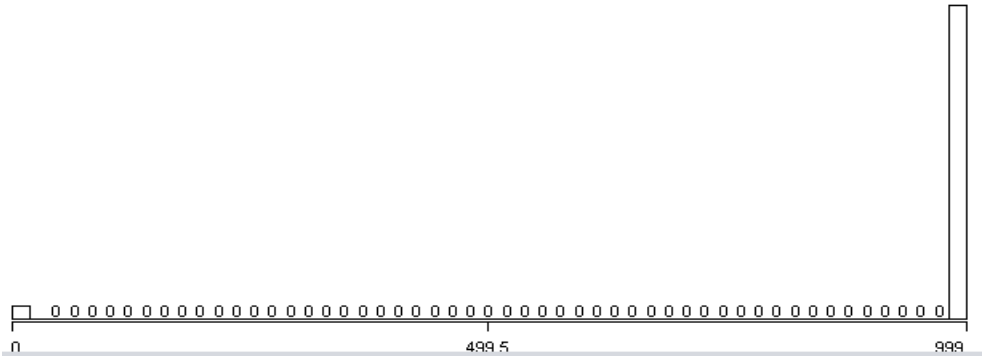


4) أيام عمل: (pdays)

عدد الأيام التي مرت بعد آخر اتصال للعميل من حملة سابقة 999 (يعني أنه لم يتم الاتصال بالعميل سابقًا) وكان التوزيع كالتالي:

Statistic	Value
Minimum	0
Maximum	999
Mean	962.475
StdDev	186.911

Class: pdays (Num) ▼ Visualize All



تعد هذه السمة من أكثر السمات التي تحتوي على نقاط شاذة ولاكن لا يجب حذفها لأنها تحمل معلومات هامة أي أنه لم يمر أيام على الاتصال بالعملاء من حملات تسويقية سابقة مما يمكن تميزه عمن قد مرت أيام عن اتصال البنك به مع أنهم قلة جدا لكن منهم من وضع وديعة فهذا ما سيتركن نتضمنه ولا نزيله مع انها تملك نقاط شاذة، وكان التقسيم كالتالي:

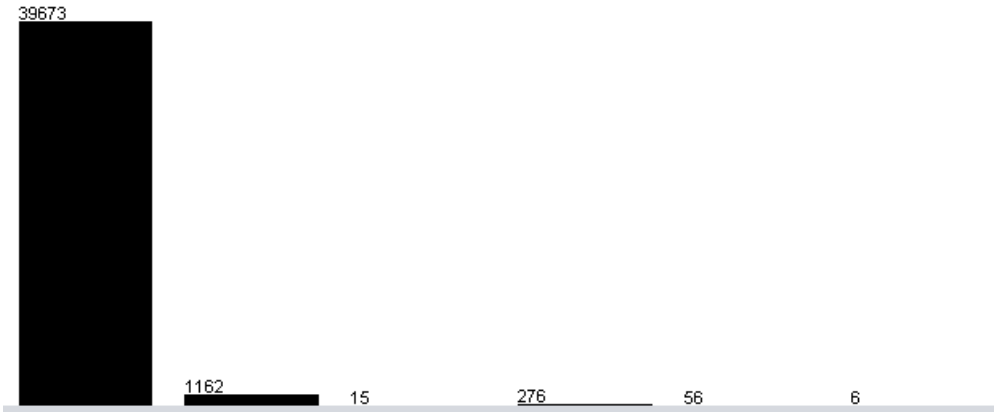
- 0 تعني (zero)
- من 1 إلى 7 (aweek)
- من 8 إلى 15 (two week)
- من 16 إلى 23 (three week)

- من 23 إلى 30 (month)
- 999 تعني (not call)

فيصبح التوزيع لهذه السمة كالتالي:

No.	Label	Count	Weight
1	not call	39673	39673.0
2	a week	1162	1162.0
3	zero	15	15.0
4	Two weeks	276	276.0
5	three weeks	56	56.0
6	Four weeks	6	6.0

No class | Visualize All

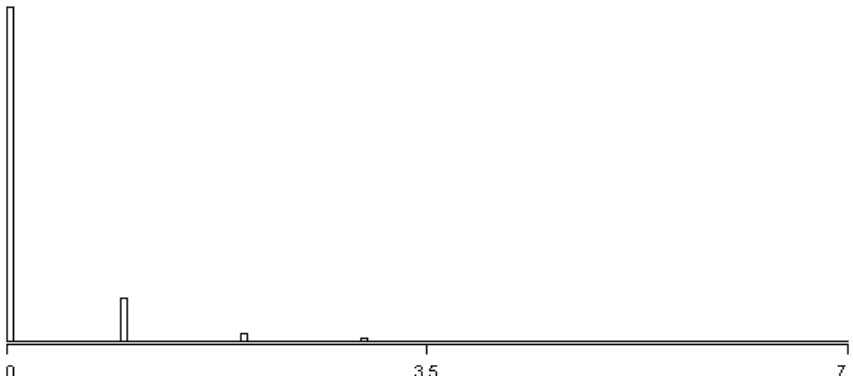


5) السابق: (Previous)

وهي عدد جهات الاتصال التي تم إجراؤها قبل هذه الحملة ولهذا العميل وكان توزيعها على الشكل التالي: حيث تتوزع القيم من 0 إلى 7

Statistic	Value
Minimum	0
Maximum	7
Mean	0.173
StdDev	0.495

Class: previous (Num) Visualize All

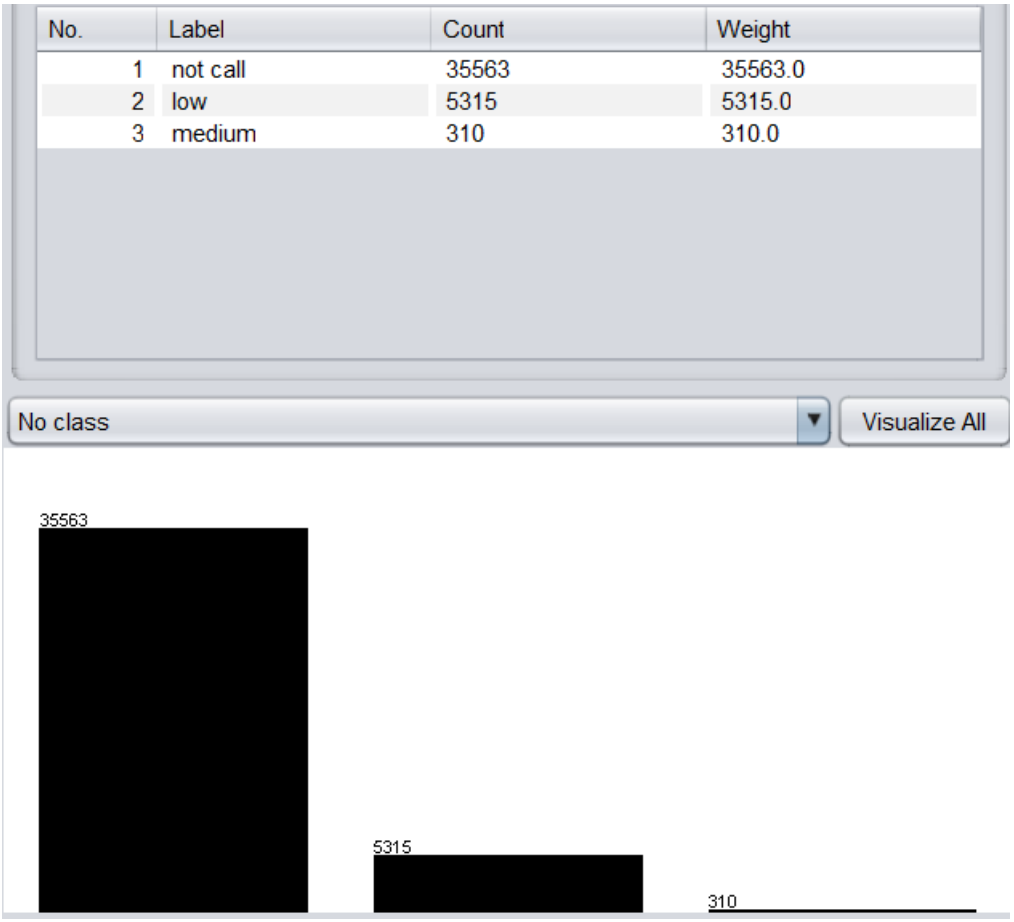


وهنا أيضا نرى أن التوزيع متشتت جدا ومن خلال الرسم البياني ويوضح ان غالب العملاء كانوا قد شملتهم الحملة مجددا إما هم عملاء جدد على البنك وقلة من هم على اتصال من الحملات السابقة، ومع ذلك فيجب الاحتفاظ بهاذه السمة على الرغم من نقاطها الشاذة الا أن المعلومات التي بها تشكل عائقا كبيرا على الحملة من دونها حتى على دقة المصنف فالعملاء الذين قد شملتهم اتصالات من حملات سابقة يتميزون عن العملاء الذين قد شملتهم الحملة لأول مرة أو إن كانوا عملاء جدد، وتتكون التقسيمات لهذه السمة كالتالي:

- 0 يعني (not call)
- من 1 إلى 2 اتصال تعني (low)
- من 3 إلى 7 اتصال تعني (med)

تنويه: وهنا أيضا لا يمكننا عند التحويل إلى شرائح بأن تكون عدد المكالمات مثلا (مكالمة ونصف) لذلك يجب أن نتقيد بنوعية البيانات المدروسة فالبيانات هي التي تحدد كيفية التقسيم.

وقد أصبح التوزيع كالتالي:



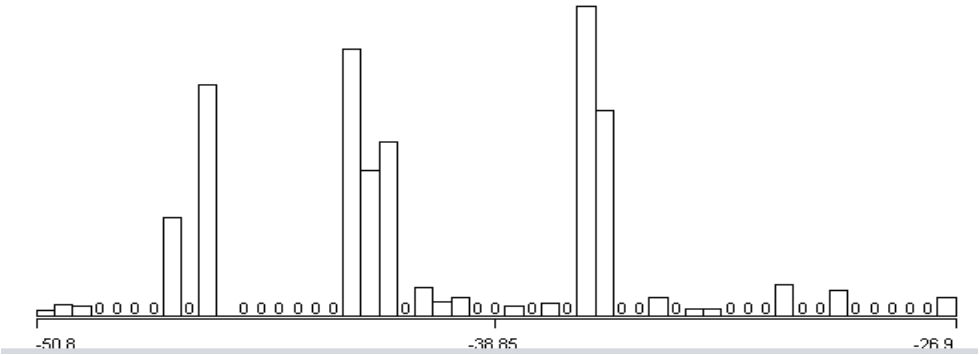
ومن هذا الرسم يتضح أن مجمل الذين شملتهم الحملة هم عملاء جدد أو أول مرة تشملهم الحملة التسويقية عما سبق، وهذا الرسم البياني يتناسب طردا مع السمة السابقة ويؤكد أن العملاء أول مرة تشملهم الحملة لأن عدد الأيام التي مرت بعد حملة سابقة كانت الغالبية لا يملكون أيام والتي كانت (999) وأصبحت بعد التحويل فئة (not call) أيضا.

6) مؤشر ثقة المستهلك: (Cons.conf.idx)

تتوزع قيم المؤشر علة النحو التالي:

Statistic	Value
Minimum	-50.8
Maximum	-26.9
Mean	-40.503
StdDev	4.628

Class: cons.conf.idx (Num) Visualize All



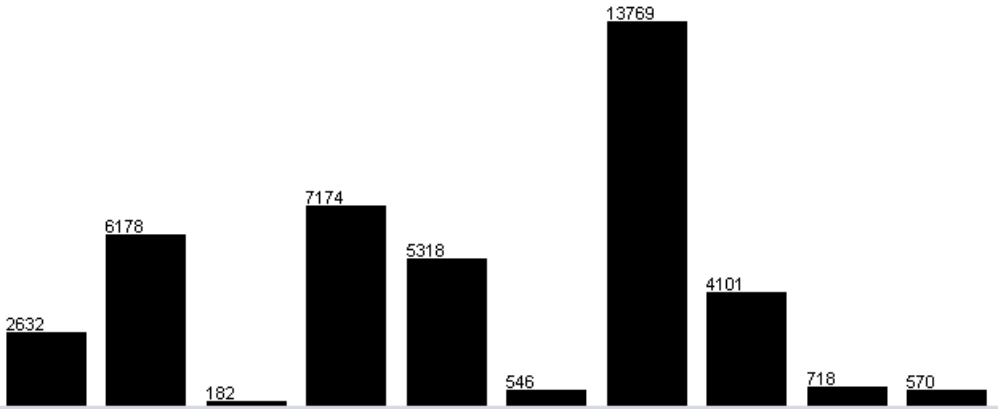
يظهر الرسم البياني التوزيع للمؤشر وكما ذكرنا سابقا إن هذا المؤشر هو مؤشر شهري ثابت لا يتغير إلا كل شهر فسيتم التقسيم على أساس الأشهر أي أن كل نشرة للمؤشر ستأخذ أسم الشهر التي تنشر فيه، وهم كالتالي:

(march / april / may / june / july / august / September / October / November / December)

وسيصبح التقسم للفئات كالتالي:

No.	Label	Count	Weight
1	apr	2632	2632.0
2	aug	6178	6178.0
3	12	182	182.0
4	jul	7174	7174.0
5	jun	5318	5318.0
6	mar	546	546.0
7	may	13769	13769.0
8	nov	4101	4101.0
9	oct	718	718.0
10	sep	570	570.0

Class: nr.employed (Num) Visualize All



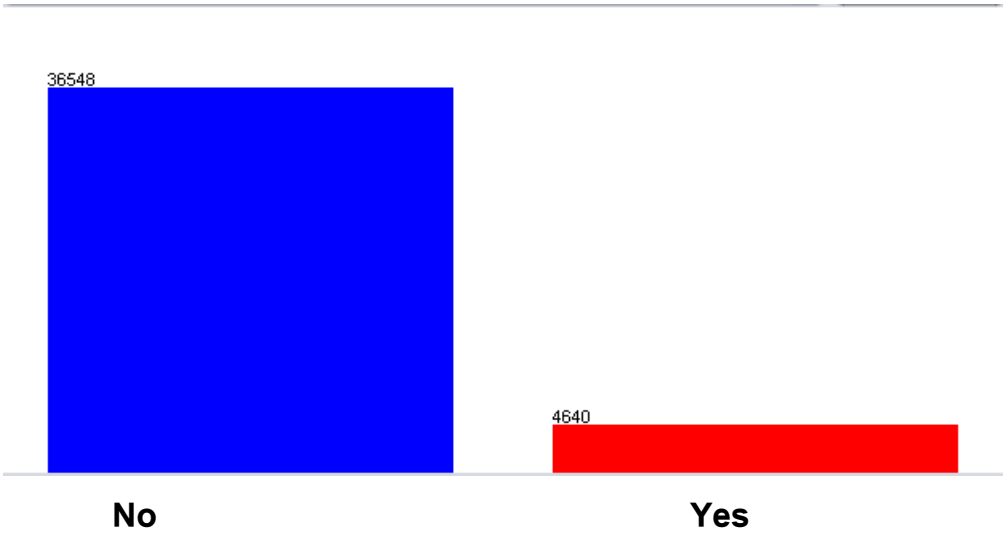
7) سعر اليوروبير 3

إن هذه السمة هي مؤشر يومي ولا يمكن وضعها ضمن فئات لذلك سيتم تضمينها كما هي

4-1-6- مشاكل التصنيف غير المتوازن

تشير البيانات غير المتوازنة عادةً إلى مشكلة تتعلق بمشاكل التصنيف حيث لا يتم تمثيل الفئات بشكل متساوٍ، وغالبا ما تكون مجموعات البيانات الغير متوازنة حالة خاصة لمشكلة التصنيف حيث لا يكون توزيع الفئة متجانساً بين الفئات التي توجد في حالات التصنيف التي تتكون من فئتين: فئة الأغلبية (no)، والأقلية (yes)، وإن كان مراد التصنيف هو لدراسة الأقلية (yes) فيجب إيجاد طريقة تحل هذه المشكلة لأن التعامل مع هذه المجموعات يعتبر مشكلة حيث ان خوارزميات التصنيف القياسية عادة ما تعتبر مجموعة التدريب متوازنة وهذا يفترض وجود تحيز تجاه طبقة الأغلبية.

وفي بياناتنا المدروسة التي تتوقع إيداعات العملاء فكانت سمة الإخراج (yes/no) غير متوازنة وكانت تميل الكفة إلى جانب العملاء التي لم تشترك بوديعة استثمارية يعني أن فئة ال (no) هي الغالبة مقارنتا بالعملاء الذين اشتركوا في وديعة (yes)، على الشكل التالي:

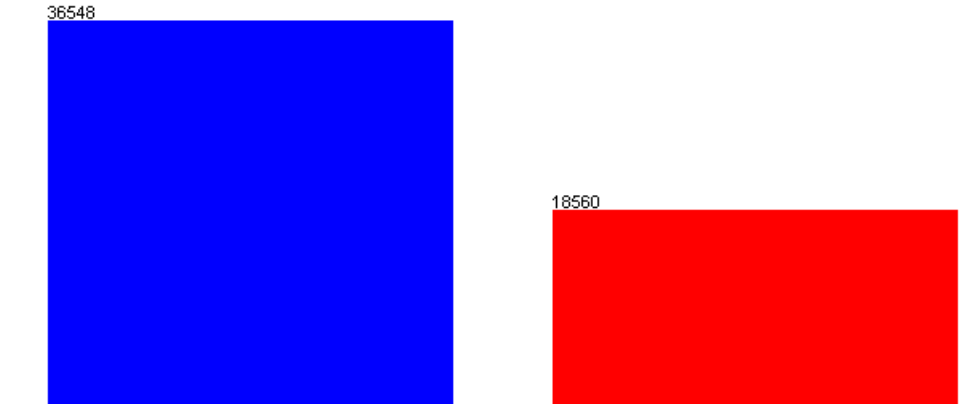


حيث كان عدد العملاء الذين اشتركوا بوديعة (4640) وقد شكلت نسبتم من البيانات الكلية (12%)، وكان عدد العملاء الذين لم يشتركوا (36548) وكانت نسبتهم من البيانات الكلية (88%)

وأن عدم التوازن في التوزيع هو واضح، وبما أن الحملة التسويقية الهاتفية طرحت لكي تبيع العملاء الودائع الاستثمارية فإن فئة الأقلية (yes) هي ذات الأهمية للحملات التسويقية.

ومن خلال التجريب على البيانات بين مفهومين (Random undersampling) الذي يعني التقليل من سجلات فئة الأغلبية (no) و (Random oversampling) والذي يعني تكرار فئة الأقلية، وبعد التجريب على البيانات اتضح إن تطبيق (Random oversampling) والتي تعني زيادة عشوائية لقيم البيانات التي تشكل الأقلية، التي تمثل في بياناتنا (yes) هي الأفضل من حيث التوازن والأفضل من خلال التطبيق على المصنفات، حيث كان أفضل توازن محقق عند تكرار سجلات فئة (yes) ثلاثة مرت.

أصبحت سمة الإخراج بعد تطبيق مفهوم (Random oversampling) كالتالي من حيث التوازن:



قد أصبح عدد السجلات الكلي (55.108) من خلال تكرار فئة (yes)، فقد أصبحت سجلات فئة (yes) تتكون من (18560) سجل بنبة من البيانات الكلية (34%)

وهكذا أصبحت البيانات متوازنة ويمكن تطبيق المصنفات عليها.

5-1-6- التصنيف: (Classification)

1-5-1-6- بناء المصنف:

من اجل تحديد دقة تصنيف البيانات التي تم تدريبها تم اعتماد أسلوب الاختبار ما يعرف (Cross validation) بنسبة تقطيع متساوية 10 وذلك لجميع حالات التصنيف في هذا البحث ويمكن تلخيص عمل هذه الطريقة بالخطوات التالية:

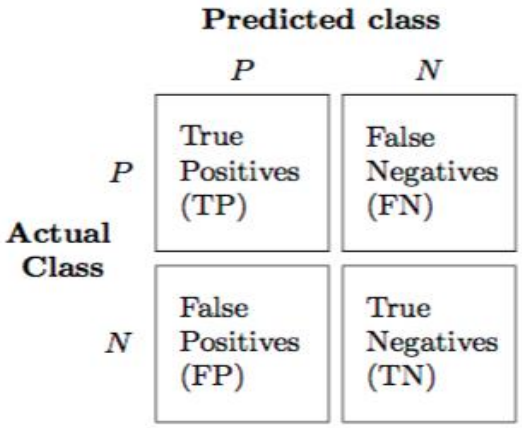
- تدريب بنسبة 100% من بيانات التدريب.
- تقسيم بيانات التدريب الى 10 اقسام بشكل عشوائي.
- تنفيذ عملية تخمين المقادير لكل من هذه الأقسام العشرة وحساب نسبة تطابق القيم الحقيقية لهذه البيانات مع القيم التي تم تخمينها.
- حساب هذه النسبة لكامل مجموعة البيانات لتكون نسبة الدقة المطلوبة.

2-5-1-6- تقييم أداء المصنف:

يتم تقيم أداء المصنفات من خلال المعلومات الناتجة من مصفوفة الارتباك (Confusion Matrix) والتي تعبر عن أداة تحليلات تنبؤية على وجه التحديد، هو جدول يعرض ويقارن القيم الفعلية مع القيم المتوقعة للنموذج في سياق التنقيب في المعطيات، ويتم استخدام مصفوفة الارتباك كمقياس لتحليل كيفية أداء المصنف على مجموعة البيانات، تولد مصفوفة الارتباك تصورًا لمقاييس مثل الدقة (accuracy) والإحكام (precision) والحساسية (recall).

السبب في أن مصفوفة الارتباك مفيدة بشكل خاص هو أنه على عكس الأنواع الأخرى من مقاييس التصنيف مثل الدقة البسيطة، فإن مصفوفة الارتباك تولد صورة أكثر اكتمالاً لكيفية أداء النموذج. يمكن أن يؤدي استخدام مقياس مثل الدقة فقط إلى موقف يكون فيه النموذج يخطئ في تحديد فئة واحدة بشكل كامل ومتسق، ولكنه يمر دون أن يلاحظه أحد لأن الأداء جيد في المتوسط. وفي الوقت نفسه، تعطي مصفوفة الارتباك مقارنة بين قيم مختلفة مثل السلبيات الكاذبة والسلبيات الحقيقية والإيجابيات الكاذبة والإيجابيات الحقيقية.

يبدو شكل مصفوفة الارتباك كل تالي:



- **الدقة: (accuracy)**

يقيس عدد مرات قيام المصنف بالتنبؤ الصحيح. إنها نسبة عدد التنبؤات الصحيحة إلى إجمالي عدد التنبؤات (عدد نقاط بيانات الاختبار).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **الإحكام: (precision)**

يخبرنا ما هي نسبة الأحداث التي صنّفناها كصف معين، في الواقع كانت تلك الفئة. إنها نسبة الإيجابيات الحقيقية لجميع الإيجابيات.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **الحساسية: (recall)**

يخبرنا ما نسبة الأحداث التي كانت في الواقع من فئة معينة تم تصنيفها من قبلنا على أنها تلك الفئة. إنها نسبة الإيجابيات الحقيقية لجميع الإيجابيات.

$$\text{Recall} = \frac{TP}{TP + FN}$$

3-5-1-6- خوارزميات التصنيف: (Classification algorithms)

من خلال الأسباب التي ذكرت سابقا لاستخدام خوارزميات التصنيف في هذه الدراسة عدا عن غيرها من باقي التقنيات والخوارزميات التابعة للتنقيب في البيانات، فحسب البيانات الموجودة وطبيعتها ووجود سمة إخراج تصنف العملاء على أساسها (yes/no)، وحسب المشكلة المدروسة التي تشير إلى التنبؤ بنتائج الحملات التسويقية (الهاتفية) لتوقع إيداعات العملاء الاستثمارية، من هنا تم اعتماد تقنيات التصنيف وخوارزمياتها لبناء مصنف يتوقع ي العملاء الجدد التي سوف تستهدفهم الحملات القادمة من اتصالات، لتصنيفهم بعد إغلاق المكالمة الهاتفية مباشرة من خلال البيانات التي اخذت منهم وتصنيفهم من سيشترك في وديعة استثمارية ومن لن يشترك.

فسيتم تطبيق 5 خوارزميات للتصنيف على البيانات الموجودة والمقارنة بين المصنفات المخرجة للوصول الى أفضل مصنف ليتم اعتماده كمصنف لحملات التسويق المباشر (الهاتفية) في البنك، سيتم تطبيق خوارزميات " الغابات العشوائية (Random Forest)، شجرة القرار (Decision Tree J48)، الانحدار اللوجيستي (logistic regression)، الجار الأقرب (KNN)، وبايز الساذج (naive bayes) "

1) مصنف (J48):

كانت المخرجات من خلال استخدام مصنف **(J48)** كالتالي:

```
Time taken to build model: 1.54 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       46812               84.9459 %
Incorrectly Classified Instances      8296               15.0541 %
Kappa statistic                          0.6612
Mean absolute error                      0.2145
Root mean squared error                  0.351
Relative absolute error                 48.0253 %
Root relative squared error             74.2731 %
Total Number of Instances            55108

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.892     0.234     0.882      0.892     0.887     0.661     0.879     0.930     no
               0.766     0.108     0.783      0.766     0.774     0.661     0.879     0.762     yes
Weighted Avg.  0.849     0.192     0.849      0.849     0.849     0.661     0.879     0.874

=== Confusion Matrix ===

     a     b   <-- classified as
 32601  3947 |     a = no
  4349 14211 |     b = yes
```

كانت نتائج المصنف كالتالي:

- سرعة بناء المصنف (1.54)
- صحة دقة المصنف (84.9%)
- خطئ دقة المصنف (15%)
- Recall بالنسبة لسجلات Yes (0.766)
- Recall بالنسبة لسجلات No (0.892)

2) الغابات العشوائية: (Random Forest)

كانت مخرجات المصنف من خلال استخدام مصنف (Random Forest) كالتالي:

```
Time taken to build model: 6.24 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       52524               95.311  %
Incorrectly Classified Instances      2584                4.689  %
Kappa statistic                          0.8984
Mean absolute error                      0.1166
Root mean squared error                  0.2011
Relative absolute error                 26.105  %
Root relative squared error             42.5432 %
Total Number of Instances            55108

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.931    0.003    0.998      0.931    0.963      0.903    0.994     0.997     no
                 0.997    0.069    0.880      0.997    0.935      0.903    0.994     0.977     yes
Weighted Avg.    0.953    0.025    0.958      0.953    0.954      0.903    0.994     0.990

=== Confusion Matrix ===

     a     b   <-- classified as
 34026  2522 |     a = no
    62 18498 |     b = yes
```

كانت نتائج المصنف كالتالي:

- سرعة بناء المصنف (6.24)
- صحة دقة المصنف (95.3%)
- خطئ دقة المصنف (4.68%)
- Recall بالنسبة لسجلات Yes (0.997)
- Recall بالنسبة لسجلات No (0.931)

4) مصنف نايف بايز: (Naïve Bayes)

كانت مخرجات المصنف من خلال استخدام مصنف (Naïve Bayes) كالتالي:

```
Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       41872               75.9817 %
Incorrectly Classified Instances     13236               24.0183 %
Kappa statistic                          0.4476
Mean absolute error                      0.268
Root mean squared error                  0.4384
Relative absolute error                 59.9964 %
Root relative squared error             92.7697 %
Total Number of Instances            55108

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.846    0.411    0.802      0.846    0.824      0.449    0.774     0.836     no
                 0.589    0.154    0.661      0.589    0.623      0.449    0.774     0.680     yes
Weighted Avg.    0.760    0.324    0.755      0.760    0.756      0.449    0.774     0.784

=== Confusion Matrix ===

     a     b   <-- classified as
 30936  5612 |     a = no
  7624 10936 |     b = yes
```

كانت نتائج المصنف كالتالي:

- سرعة بناء المصنف (0.06)
- صحة دقة المصنف (75.9%)
- خطئ دقة المصنف (24%)
- Recall بالنسبة لسجلات Yes (0.589)
- Recall بالنسبة لسجلات No (0.846)

4) الانحدار اللوجستي: (logistic regression)

كانت مخرجات المصنف من خلال استخدام مصنف (logistic regression) كالتالي:

```
Time taken to build model: 5.99 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       43404               78.7617 %
Incorrectly Classified Instances     11704               21.2383 %
Kappa statistic                          0.4894
Mean absolute error                      0.3211
Root mean squared error                  0.4001
Relative absolute error                 71.8698 %
Root relative squared error             84.6613 %
Total Number of Instances            55108

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.911    0.455    0.798      0.911    0.851      0.502    0.791     0.849     no
                 0.545    0.089    0.756      0.545    0.633      0.502    0.791     0.705     yes
Weighted Avg.    0.788    0.332    0.784      0.788    0.777      0.502    0.791     0.800

=== Confusion Matrix ===

     a     b   <-- classified as
 33292  3256 |     a = no
  8448 10112 |     b = yes
```

كانت نتائج المصنف كالتالي:

- سرعة بناء المصنف (5.99)
- صحة دقة المصنف (78.7%)
- خطئ دقة المصنف (21.2%)
- Recall بالنسبة لسجلات Yes (0.545)
- Recall بالنسبة لسجلات No (0.911)

5) مصنف الجار الأقرب: (KNN)

كانت مخرجات المصنف من خلال استخدام مصنف (KNN) كالتالي:

```
Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances       51023               92.5873 %
Incorrectly Classified Instances      4085                7.4127 %
Kappa statistic                          0.8422
Mean absolute error                      0.0754
Root mean squared error                  0.2495
Relative absolute error                 16.8712 %
Root relative squared error             52.7959 %
Total Number of Instances            55108

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.891    0.005    0.997      0.891    0.941      0.852    0.944     0.979     no
                 0.995    0.109    0.822      0.995    0.900      0.852    0.944     0.824     yes
Weighted Avg.    0.926    0.040    0.938      0.926    0.927      0.852    0.944     0.927

=== Confusion Matrix ===

     a     b   <-- classified as
 32552  3996 |     a = no
    89 18471 |     b = yes
```

كانت نتائج المصنف كالتالي:

- سرعة بناء المصنف (0.01)
- صحة دقة المصنف (92.5%)
- خطئ دقة المصنف (7.4%)
- Recall بالنسبة لسجلات Yes (0.955)
- Recall بالنسبة لسجلات No (0.891)

3-5-1-6- اختيار المصنف

قد تم من خلال بناء المصنفات الخمس السابقة " الغابات العشوائية (Random Forest)، شجرة القرار (Decision Tree J48)، الانحدار اللوجيستي (logistic regression)، الجار الأقرب (KNN)، وبايز الساذج (naive bayes) " تم الحصول على النتائج ومن ثم سيتم اختيار المصنف الذي يعطي أعلى دقة من خلال مقارنة المعايير التالية: (Accuracy) و(Recall) و(Precision)

criterions / name of classifier	Accuracy	Recall yes	Recall No	Precision Yes	Precision No
J48	84.9%	0.892	0.86	0.783	0.882
Random Forrest	95.3%	0.997	0.931	0.880	0.998
Naïve bayes	75.9%	0.589	0.846	0.611	0.802
logistic regression	78.7%	0.545	0.911	0.765	0.798
KNN	92.5%	0.955	0.891	0.822	0.997

من خلال المقارنة بين نتائج المصنفات ال خمس المستخدمة أتضح أن مصنف الغابات العشوائية (Random Forrest) قد حصل تميز بجميع نتائج معايره على المعايير الأخرى من حيث الدقة (accuracy) والإحكام (precision) والحساسية (recall)، ويليه المصنف الجار الأقرب (KNN)، مجملا أن مصنف (Random Forrest) كان مرشحا جدا لأنه غالبًا ما يكون مصنفًا دقيقًا تمامًا ويعمل بشكل جيد مع الميزات الثنائية ومجموعات البيانات الكبيرة.

الفصل السابع

النتائج والتوصيات والمراجع

يهدف هذا الفصل إلى عرض عن أهم النتائج التي توصلت إليه هذه الدراسة من الجانب النظري والجانب التطبيقي التي توصل اليها الباحث، كإجابة عن الأسئلة التي تم طرحها في الفصل الاول من الدراسة والتي مثلت مشكلتها، وفي ضوء النتائج التي توصلت اليها الدراسة قدم الباحث مجموعة من التوصيات

1-7- النتائج:

من خلال الدراسة التي أجريت على بيانات البنك البرتغالي (Banco Português de Investimento) التي تخص استخدام تقنيات التنقيب في المعطيات للتنبؤ بنتيجة الحملات التسويقية المباشرة (الهاتفية) لتوقع إيداعات العملاء البنكية من خلال ما تم توضيحه في الشقين النظري والعملي التطبيقي من البحث يمكن تلخيص النتائج التي توصل لها الباحث من خلال بحثه بما يلي:

- **النتائج النظرية للدراسة:**

1) الاستخدامات العديدة التي تقدمها علوم البيان من خلال التنقيب في المعطيات وتقنياتها العديدة في مجال البنوك بشكل عام ومن ثم تطبيق التقنيات على الأقسام العديدة التابعة للبنك (قسم التسويق) بشكل خاص، لما له محاسن من خلال التنبؤ ووصف العميل التابع للبنك بشكل دقيق عبر البيانات الخاصة بالعميل.
2) لم يعد للمدراء المقدرة على قياس تلك البيانات من خلال النظر اليها، ومن جوهر المشكلة تأتي النتيجة باستخدام تقنيات التنقيب في المعطيات لفهم العميل واحتياجاته وتصنيفه والتنبؤ بالحركة القادمة له، مما يعطي مصداقية أكبر للعميل اتجاه الحملات وفي المقابل يحصر البيانات اللازمة للحملة من خلال هندسة البيانات، مما يحدد البيانات اللازمة معرفتها من العميل دون تطفل وإزعاج لنجاح نتائج الحملات التسويقية على أتم وجه.
3) تطبيق تقنيات التنقيب في المعطيات (التصنيف) لتوقع إيداعات العملاء يتيح النظر مرة أخرى في البيانات لهندستها من بعد تحديد قسم التسويق للبيانات اللازمة من خلال تحضير البيانات الخاصة بالحملة التسويقية ومعرفة تأثيرات وارتباطات سمات الحملة بين بعضها وحذف السمات التي قد تؤثر على دقة التنبؤ.
4) استخدام خوارزميات التصنيف في المشكلة المدروسة وعدم استخدام تقنيات أخرى يعطي أفضل نتائج من خلال طبيعة البيانات الموجودة، وعلى أساسها أصبح ممكنا التوقع لإيداعات العملاء بشكل كبير والنتائج التطبيقية أثبتت ذلك.
5) إضاقة السمات الاقتصادية والاجتماعية أعطت التنبؤ بنتائج الحملات التسويقية المباشرة دقة كبيرة بتوقع إيداعات العملاء، لأنها تربط بين ثقافة وثقة ورأي مجتمع كامل من جهة وتغير الأسعار وزيادة وانخفاض أسعار الفائدة على القروض من جهة أخرى على

الصعيد المحلي البرتغالي، مما يربط بين الأموال التي يملكونها العملاء وإمكانية إيداعاتهم من خلال تحليل المؤشرات الصادرة من البنك البرتغالي المركزي.

6) استخدام تقنيات التنقيب في البيانات لها أثر كبير على تخفيض تكلفة الحملات التسويقية وإدارة الحملة بالشكل الأمثل، وزيادة إيرادات البنك عن طريق التوقع الصحيح للعملاء الذين سيشتركون بودائع استثمارية، واستهداف شرائح كانت مهمشة وغير مركز عليها.

- **النتائج التطبيقية للدراسة:**

من خلال الدراسة العملية والتحليلية لبيانات هذا البحث كانت النتائج كالتالي:

1) يعرف بأن محور الحملة التسويقية المباشرة (الهاتفية) هي المكالمة لذلك يمكن من خلال الالتزام بمدة المكالمة التي حددت وهي بين (9) و (14) دقيقة فيمكن الحصول على إيداعات من العملاء بنسبة أكبر وبالتالي تخفيض تكلفة الحملات من خلال المعرفة التقريبية لنسبة مدة الاتصال.

2) يجب التركيز على الفئات العمرية التي بين (30) و (45) لأنها تملك أكبر شريحة عمرية من العملاء بشكل عام، وبشكل خاص هم أكثر شريحة تشترك بودائع بين العملاء المستهدفين، الاستفادة من الفئة العمرية ما فوق (60) بسبب حركة تجاوبهم مع الحملة التسويقية لكن بنسب صغيرة.

3) يمكن للمؤشرين سعر يوربيور (يوريبور 3) وثقة المستهلك (cons.conf.idx) عند معدلات معينة في أيام وأشهر إمكانية الاعتماد عليهم بالنسبة لتكثيف الحملة مرتبطا بمدة المكالمة الناتجة للحصول على حركة إيداع أكبر من قبل العملاء، وتخفيض الحملة نسبيا بالمعدلات التي أظهر فيها نسب كثيرة لعدم الإيداع، حيث المؤشرين يؤثران بشكل ملحوظ بعملية نتائج الحملة وهذا منطقي من خلال المنطق الاقتصادي بحيث هناك ارتباط بين قدرة العميل على الشراء ورأيه وأمانه على نسبة إيداعه، وكذلك سعر اليوربيور بحث معرفته بسعر الفائدة على معاملات اليورو ما بين المصارف يأثر علة احتمالية إيداعه.

4) تم بناء (5) مصنفات للبيانات المدروسة واختيار أمثل مصنف واعتماده لتوقع إيداعات العملاء الاستثمارية، وقد كان مصنف الغابات العشوائية هو الأفضل من بين المصنفات المستخدمة متفوقا عليهم بجميع المعاير بنسبة دقة (95.3%)

5) يجب الاخذ بعين الاعتبار جميع النتائج بالإضافة لمدة المكالمة مع للجميع فهي أساس نجاح الحملة إن أتمت على أكمل وجه مع النتائج السابقة.

2-7- التوصيات

- **توصية خاصة للبنوك**

- ❖ تطبيق تقنيات التنقيب في البيانات على جميع أقسام البنك وعدم اقتصار قسم معين (تسويق) لما له من ميزات على المدى الطويل والارتباط القوي داخل البنك وتعامله مع العميل.
- ❖ إمكانية تحسين الحملات التسويقية المباشرة (الهاتفية) من خلال اختيار تقنية العنقدة لتجميع العملاء على عدة اقسام والتعامل مع قل قسم بحملة تسويقية خاصة لما له مزايا عديدة على نطاق الحملة التسويقية من تخفيض الجهد والمدة والتكلفة، وعلى نطاق البنك من خلال زيادة الإيرادات.
- ❖ التنبؤ المسبق لحاجات العميل ورغباته من أجل اشباعها يكمن باستخدام تقنيات التنقيب في البيانات.
- ❖ عندما يتم إقرار البيانات التي يجب الحصول عليها من العملاء يجب أن تكون مدروسة بشكل صحيح لتعطي معلومات مفيدة ة والتخلص من البيانات التي لا قيمة لها لكي يتم الاستخدام الأمثل لها من خلال التقنيات العديدة للتنقيب في البيانات.
- ❖ فهم المشكلة المراد استخدام تقنيات التنقيب في البيانات عليها لتحديد التقنية المناسبة لها والحصول على النتائج الأمثل (تنبؤ /قواعد ترافق/ تصنيف)
- ❖ اقتصرت الدراسة على بيانات قليلة نسبيا خاصة بالعميل فيجب على إدارة الحملات التسويقية للبنوك الصول لمعلومات عن العميل تفيد بالشكل المناسب.
- ❖ إمكانية التنبؤ بسلوك العميل الذي يصعب التنبؤ فيه مجملا، عن طريق تقنيات التنقيب التي تندرج تحت أطار التنبؤ
- ❖ العمل على دراسة للتنبؤ بإقراض البنوك للعملاء من خلال بياناتهم
- ❖ استخدام تقنيات التنقيب في البيانات على جميع الأصعدة تزيد من ايرادات البنوك وتسهيل عملياتها الداخلية والخارجة والاحتكاك مع العميل بكل وقت.

- **دراسات خاصة للاكاديميين والدراسات المستقبلية**

❖ التوسع في هذه الدارسة من خلال الاستفادة من خوارزميات أخرى للتنقيب في المعطيات لمعالجة هذا النوع من المعطيات وعلى وجوه الخصوص استخدام خوارزميات أخرى غير التصنيف مثل العنقدة وقواعد الترافق سعياً للوصول الى نتائج أفضل.

❖ استخدام أدوات مختلفة عن الذي اعتمدها البحث لتقنيات التنقيب في البيانات.

❖ تطبيق تقنيات التنقيب في البيانات على جميع المجالات لما لها من ميزات وفائدة مدروسة ومحققة للجهة المطبق عليها.

❖ فهم جميع البيانات بشكل دقيق وانتقاء التقنية الأنسب.

❖ أن يكون هذا البحث مرجع غني لباقي الدراسات المستقبلية

1) مراجع

2) *Most Common Applications and Techniques of Data Mining* . (2020).

3) Piateski, G., & Frawley, W. (1991). *Knowledge Discovery in Databases.* Cambridge, MA.

4) Biao Qin ،Yuni Xia ،Shan Wang و ،Xiaoyong Du. .(بلا تاريخ) *A Novel Bayesian Classification Technique for Uncertain Data.* China: University Indianapolis.

5) Breiman, L. (September 2010). *Random Forests for Regression and classification.* Ovronnaz, Switzerland: Utah State University.

6) han, j., kamber, n., & pei, j. (2000). *Data Mining concepts and techniques.*

7) Harpale, A. S. (2003). *Clustering.* Bombay. powai: Indian Institute of technology.

8) *https://www.bancobpi.pt/grupo-bpi/grupo-bpi-grupo-bpi*. (n.d.).

9) Imandoust, Bafandeh, S., & Bolandraftar. (2013). *Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background.* Journal of Engineering Research and Applications.

10) Jared dean. .(2014) *Big Data, Data Mining, and Machine Learning.* wiley.

11) Lu, D. M. (n.d.). *Data warehousing and Data Mining.*

12) Lu, D. M. (n.d.). *Data warehousing and Data Mining* .

13) Morgan, S., & Teachman, J. (nov/1988). *Logistic Regression: Description, Examples, and Comparisons.* Journal of Marriage and Family.

14) Palace, B. (1996). *Data Mining, Technology note prepared for Management.* UCLA.

15) Prediction, D. M.-C. (n.d.). Great Islamians.

16) Raschka, S. (October 4, 2014). *Naive Bayes and Text Classification.*

17) romero. (n.d.). *Data Mining and Data –Classification- the classificatoin problem.*

18) sejdinovic, D. (n.d.). *Statistical Data Mining and Machine Learning.* Oxford.

19) Veldhoen, A., & De Prins, S. (2004). *Applying Big Data To Risk Management.*

20) Witten , I., & Frank, E. (2005). *Practical Machine Learning Tools and Techniques.*

21) *www.cs.waikato.ac.nz*. (n.d.).

22) ابتهال فارس جاسم الحسون. (2016). *اهمية واهداف التسويق المصرفي وأثره على كفائة المصارف*. جامعة بابل للعلوم.

23) احمد البغى. (2019). *التسويق المصرقي*.

24) الموقع الرسمي للبنك. (بلا تاريخ). *www.bancobpi.pt/grupo-bpi/grupo-bpi-grupo-bpi*.

25) جلال الضاهر. (2014). *تصميم نموذج نظام دعم القرار إدارة الموارد البشرية بالعتماد عمى تقنيات الذكاء الصنعي*. الجامعة الافتراضية السورية.

26) خير الدين محمد، و عبد داود. (2012). *Development of important patterns in student database using statistical classifier.* جامعة الأنبار.

27) د. باسل يوسف، عبدالله يوسف سعادة، و هيام محمد الزيدانيين. (2016). *أثر توظيف الودائع الاستسمارية في ربحية المساهمين*. الجامعة الاردنية.

28) زكريا مطللك، و داليا عبد الحسن. (بلا تاريخ). *دور التنقيب في البيانات في زيادة أداء المظمة*. 2007: جامعة بغداد.

29) سارة بن ذيب، و راضية نصر الله. (2016). *التسويق المصرفي وتحديات تكنلوجيا الاعلام والاتصال*. جامعة العربي التبسي.

30) سعدون حمود جسير. (2018). *دراسة أساليب التسويق المباشر وتحديد مدى ملائمة كل منها للتطبيق في السوق العراقية*. جامعة بغداد.

31) سفيان مسالتة. (2017). *دور سياسات التسويق المباشر في تحيسن أداء الخدمات وتحقيق الميزة التنافسية*. جامعة سطيف.

32) سمية فجخي. (2017). *أثر التسويق المصرفي في تحقيق الميزة التنافسية للبنوك*. جامعة أم البواقي.

33) عبيد, م .*data analysis and data mining*. (2019). دار الفكر العربي / القاهرة.

34) غربي يسين الاخضر. (2013). *دور إدارة الودائع في التوفيك بين هدفين السيولة والربحية لدى البنوك الجزائرية*. جامعة الجلفة.

35) للبنك, ا. ا.*www.bancobpi.pt/grupo-bpi/grupo-bpi-grupo-bpi*. (n.d.) .

36) مايك تشابل. (بلا تاريخ). *تقنيات التصنيف تحليل البيانات والتنبؤ*.

37) نجيب سهيل سلق. (بلا تاريخ). *بناء نظام مساعد في اتخاذ القرار لحل مشكلة تسرب التالميذ في مرحلة التعليم االلزامي*. دمشق: المعهد العالي للعلوم التطبيقية والتكنلوجيا.

38) نور عباس عمران النصراوي. (2017). *استخدام أسلوب البوتوستراب في تحليل النماذج المعلمية وشبه المعلمية والمقارنة بينهما*. جامعة كربلاء.