وزارة التعليم العالـي والبحث العلمي

جامعة العربي التبسي -تبسة-

كلية العلوم الاقتصادية، العلوم التجارية وعلوم التسيير

قسم: علوم التسيير الرقم التسلسلي:...../2026

مذكرة مقدمة لاستكمال متطلبات شهادة ماستر أكاديمي (ل م د)

دفعـــــة: 2026

فرع: علوم التسيير

التخصص: إدارة أعمال

تصميم نموذج تنبؤي لسلوك الموظفين باستخدام تقنيات تعلم الآلة
حالة تسرب الموظفين باستعمال قاعدة بيانات لشركة IBM

من إعداد الطالبين:
- عزاز وائل
- قريب صهيب

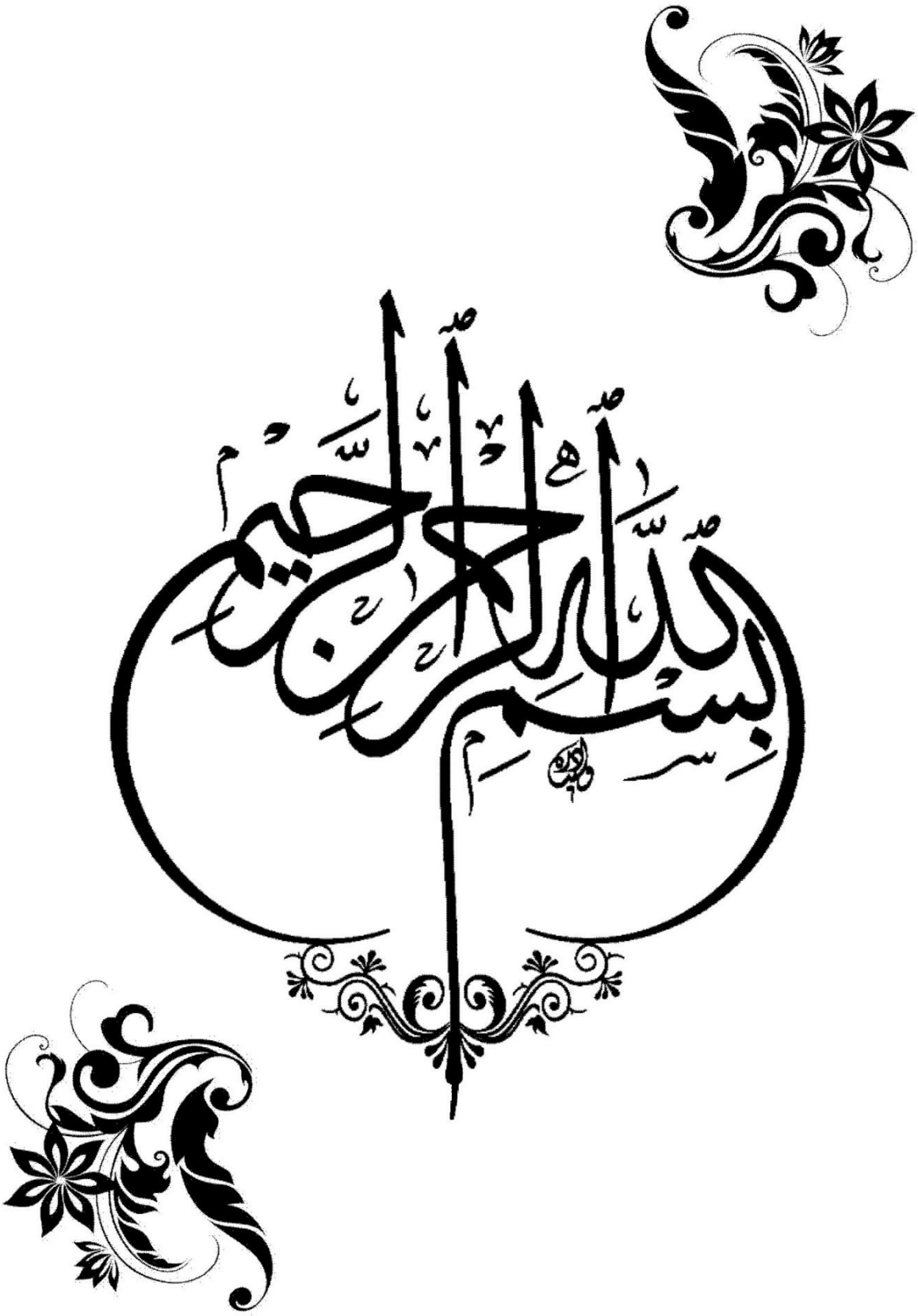
تحت إشراف الدكتور
- حميدان ربيع

جامعة العربي التبسي - تبسة
Université Larbi Tebessi - Tebessa

لجنة المناقشة

الاسم واللقب	الرتبة العلمية	الصفة
يوسف عبايدية	أستاذ محاضر - أ-	رئيسا
حميدان ربيع	أستاذ محاضر - ب-	مشرفا ومقررا
عبد الكريم شوكال	أستاذ محاضر - ب-	مناقشا

السنة الجامعية 2026/2025


بسم الله الرحمن الرحيم

شكر وعرفان

الحمد لله حمدًا كثيرًا طيبًا مباركًا فيه حتى يبلغ الحمد منتهاه، والصلاة والسلام على أشرف الخلق وسيد المرسلين سيدنا محمد صلى الله عليه وسلم، أما بعد:

قال رسول الله صلى الله عليه وسلم :"من لم يشكر الناس لم يشكر الله"، وانطلاقًا من هذا المبدأ، لا يسعني إلا أن أتقدم بخالص الشكر والتقدير وعظيم الامتنان إلى أستاذي الفاضل حميدان ربيع الذي تفضل بالإشراف على هذا العمل، وكان سندًا وداعمًا في مختلف مراحل إنجازه، من خلال توجيهاته القيّمة وإرشاداته السديدة، وما قدمه من مساعدة ونصح لإتمام هذا العمل المتواضع.

ولا يفوتني أن أتقدم بخالص الشكر والامتنان إلى جميع الأساتذة الكرام الذين أشرفوا على تدريسي طوال مشواري الجامعي، فقد كانت لبصماتهم وآثارهم العلمية الدور الكبير في وصولي إلى هذه المرحلة، فلكم مني كل الاحترام والتقدير.

والحمد لله الذي يسّر لي سنوات الدراسة، وأكرمني بنعمة التعلم والتوجيه على أيدي أساتذة أفاضل، أسأل الله أن يبارك في علمهم، ويجزيهم خير الجزاء.

وفي الختام، أرجو من الله تعالى أن يجعل هذا العمل صدقة جارية، ينتفع بها كل من أراد البحث أو الاطلاع على موضوع هذه المذكرة، وأن يكتب له القبول والنفع.

الفهرس العام

العنوان	الصفحة
المقدمة	1
الفصل الأول: الإطار النظري	
تمهيد	4
المبحث الأول: مدخل الى سلوك الموظف داخل المنظمة	
المطلب الأول: مفهوم سلوك الموظف	4
الفرع الأول: تعريف سلوك الموظف	4
الفرع الثاني: أهمية دراسة سلوك الموظف في المؤسسة	5
المطلب الثاني: محددات سلوك الموظف	6
الفرع الأول: المحددات الفردية	6
الفرع الثاني: المحددات التنظيمية	7
المطلب الثالث: سلوكيات الموظفين القابلة للتنبؤ	8
المبحث الثاني: عموميات حول التنبؤ و تقنيات تعلم الآلة	
المطلب الأول: عموميات حول التنبؤ	10
أولا: تعريف التنبؤ	10
ثانيا: أهمية التنبؤ	10
ثالثا: مراحل التنبؤ	11
رابعا: طرق التنبؤ	12
المطلب الثاني: مفاهيم تعلم الآلة	14
اولا: تعريف تعلم الآلة	15
ثانيا: انواع التعلم الآلي	15
المطلب الثالث: العلاقة بين تعلم الآلة و ادارة الموارد البشرية	23
اولا: استخدامات تعلم الآلة في ادارة الموارد البشرية	23
ثانيا: تحديات استخدام تعلم الآلة في ادارة الموارد البشرية	24
المبحث الثالث: الدراسات السابقة	
المطلب الأول: الدراسات العربية	26
الدراسة الأولى	26

الدراسة الثانية	26
الدراسة الثالثة	26
المطلب الثاني: الدراسات الاجنبية	27
الدراسة الأولى	27
الدراسة الثانية	27
الدراسة الثالثة	28
الدراسة الرابعة	28
الدراسة الخامسة	29
المطلب الثالث: التعقيب على الدراسات السابقة	30
خلاصة الفصل	30
الفصل الثاني: الفصل التطبيقي	
تمهيد	32
المبحث الأول: إستكشاف بيانات الدراسة	
المطلب الأول: التعريف بمتغيرات الدراسة ومصدر جمع البيانات	32
أولا : مصدر البيانات	32
ثانيا : وصف البيانات	33
المطلب الثاني: أدوات وبيئة التطوير المستخدمة في الدراسة	34
أولا: برمجة Python	34
ثانيا: بيئة التطوير (Google Colab)	36
المطلب الثالث: خطوات بناء النموذج وأدوات تقييمه	36
اولا :مراحل بناء النموذج	36
ثانيا: أدوات تقييم النموذج	39
المبحث الثاني: تحليل النتائج وتفسيرها	
المطلب الأول: استكشاف البيانات و معالجتها	43
أولا: إستيراد المكتبات الأساسية الخاصة بالبرمجة	43
ثانيا : إستيراد قاعدة البيانات	43
ثالثا : تحليل إستكشاف البيانات	44

رابعا: تقسيم البيانات	46
خامسا: تحجيم البيانات و ترميزها	46
سادسا: موازنة البيانات	46
سابعا: تحديد المتغيرات	48
المطلب الثاني: بناء و تقييم النموذج	50
أولا : شجرة القرار Decision Tree	50
ثانيا : الإنحدار اللوجستي	54
ثالثا : الغابة العشوائية	58
رابعا : التدرج المعزز	61
خامسا : الة الدعم الشعاعي	65
المطلب الثالث: المقارنة بين الدراسة الحالية و الدراسة السابقة	70
خلاصة الفصل	70
الخاتمة	72
قائمة المراجع	74
ملخص الدراسة	79

قائمة الجداول

الرقم	العنوان	الصفحة
01	المتغيرات المعتمد عليها في الدراسة	33
02	مصفوفة الالتباس Counfusion Matrix	39
03	بعض عينات قاعدة البيانات بإستخدام الأمر df.head()	44
04	المعلومات الإحصائية لقاعدة البيانات قبل المعالجة	45
05	نتائج مصفوفة الارتباك الخاصة بشجرة القرار	51
06	المؤشرات العامة حسب نموذج شجرة القرار	51
07	نتائج مصفوفة التحقق المتقاطع لنموذج شجرة القرار	52
08	نسبة مساهمة المتغيرات في بناء نموذج شجرة القرار	53
09	نتائج مصفوفة الارتباك الخاصة بالانحدار اللوجستي	55
10	المؤشرات العامة حسب نموذج الانحدار اللوجستي	56
11	نتائج مصفوفة التحقق المتقاطع لنموذج الانحدار اللوجستي	56
12	نسبة مساهمة المتغيرات في بناء نموذج الانحدار اللوجستي	57
13	نتائج مصفوفة الارتباك الخاصة بالغابة العشوائية	59
14	المؤشرات العامة حسب نموذج الغابة العشوائية	59
15	نتائج مصفوفة التحقق المتقاطع لنموذج الغابة العشوائية	60
16	نسبة مساهمة المتغيرات في بناء نموذج الغابة العشوائية	61
17	نتائج مصفوفة الارتباك الخاصة بالتدرج المعزز	62
18	المؤشرات العامة حسب نموذج التدرج المعزز	63
19	نتائج مصفوفة التحقق المتقاطع لنموذج التدرج المعزز	64
20	نسبة مساهمة المتغيرات في بناء نموذج التدرج المعزز	65
21	نتائج مصفوفة الارتباك الخاصة بالة الدعم الشعاعي	66
22	المؤشرات العامة حسب نموذج الة الدعم الشعاعي	67
23	نتائج مصفوفة التحقق المتقاطع لنموذج الة الدعم الشعاعي	68
24	نسبة مساهمة المتغيرات في بناء نموذج الة الدعم الشعاعي	69
25	المقارنة بين الدراسة الحالية و الدراسة السابقة	70

قائمة الأشكال

الرقم	العنوان	الصفحة
01	التصنيف حسب التعلم الخاضع للإشراف	16
02	تحليل الإنحدار حسب التعلم الخاضع للإشراف	17
03	التجميع العنقودي حسب التعلم الغير خاضع للإشراف	21
04	خطوات بناء النموذج	37
05	مصفوفة التحقق المتقاطع **Cross Validation**	42
06	استيراد المكتبات	43
07	استيراد قاعدة البيانات	43
08	امر استكشاف البيانات	44
09	تقسيم البيانات	46
10	البيانات قبل الموازنة	47
11	البيانات بعد الموازنة	47
12	استرداد نموذج شجرة القرار	50
13	أمر مصفوفة الالتباس الخاصة بنموذج شجرة القرار	50
14	استيراد نموذج الانحدار اللوجستي	54
15	أمر مصفوفة الالتباس الخاصة بنموذج الانحدار اللوجستي	55
16	استيراد نموذج الغابة العشوائية	58
17	أمر مصفوفة الالتباس الخاصة بنموذج الغابة العشوائية	58
18	استيراد نموذج التدرج المعزز	62
19	أمر مصفوفة الالتباس الخاصة بنموذج التدرج المعزز	62
20	استيراد نموذج الة الدعم الشعاعي	66
21	أمر مصفوفة الالتباس الخاصة بنموذج الة الدعم الشعاعي	66

المقدمة

المقدمة:

في ظل التحولات الجذرية التي تشهدها مجالات إدارة الأعمال والتكنولوجيا، لم يعد النجاح مقتصرا على تحليل الأداء الماضي والحالي للمؤسسة، بل بات مرتبطًا بشكل وثيق بالقدرة على صياغة خطط استباقية تستند إلى فهم محددات سلوك الموارد البشرية او الموظفين، وتتسابق الشركات اليوم لفك شفرات السلوك الوظيفي بهدف رفع كفاءة الإدارة والاستجابة السريعة للمتغيرات، سعيًا للحفاظ على مورد بشري كفؤ، إذ أصبحت جودة الخدمات والمنتجات انعكاسًا مباشرًا لجودة وكفاءة المورد البشري الذي تمتلكه المؤسسة، من هذا المنطلق، اصبح فهم سلوك الموظف مفتاحًا استراتيجيًا للنجاح، فالموظف ليس مجرد شخص يقوم بهمام محددة، بل هو فرد يتسم بسلوكيات معقدة تتأثر بدوافع ذات أبعاد متعددة (فردية، تنظيمية، وثقافية) تشكل قراراته اليومية. لذلك يتجاوز الامر رصد أداء الموظف المهني بل يصل إلى عمق العملية التفاعلية التي يتخذ من خلالها قرارات تؤثر إيجابًا أو سلبا على كيان المؤسسة، ويعد هذا الفهم مهما جدا لبناء سياسات المؤسسة، بدءا من تصميم أنظمة الرواتب وتوفير بيئة العمل الملائمة، وصولا إلى تحديد ساعات العمل المحفزة. ومن خلال تحليل الأنماط والخصائص المميزة للموظفين مثل السن، الرضا الوظيفي، مستوى الدخل و سنوات الخبرة، تتمكن الشركات من التنبؤ بالسلوكيات المستقبلية عبر تقنيات تعلم الآلة وبناء النماذج التنبؤية، مما يمنحها ميزة تنافسية تعزز من فرص استدامتها ونجاحها.

1- الإشكالية الرئيسية: فيما يتمثل افضل نموذج تنبؤي في تصنيف سلوك الموظفين؟

2- الأسئلة الفرعية:

- ما هو مفهوم سلوك الموظفين وما هي أهم تقنيات تعلم الآلة المستخدمة في التنبؤ؟
- ماهي مراحل بناء نموذج التنبؤ لتعلم الالة؟
- هل يمكن لنموذج SVM و التدرج المعزز ان يتفوق ادائهم على أداء نموذج Decision Tree و Logistic Regression و Random Forest في مهمة تصنيف سلوك الموظفين؟

3- الفرضية الرئيسية:

يعتبر نموذج التدرج المعزز GBM افضل نموذج تنبؤي في تصنيف سلوك الموظفين حيث يفوق أداء نموذج Decision Tree و نموذج Logistic Regression و Random Forest وآلة الدعم الشعاعي SVM في مهمة تصنيف سلوك الموظفين.

4- الفرضيات الفرعية:

- تعمل تقنية Cross Validation على تحسين مصداقية نتائج التقييم و يجعل النموذج قابل للتعميم و يمنع من التحيز في البيانات.
- نموذج GBM يحقق أعلى اداء مقارنة بباقي النماذج المستخدمة في مهمة تصنيف الموظفين حسب سلوكهم الوظيفي من حيث مؤشرات الدقة (Accuracy).

5-أهمية الدراسة:

- تساعد هذه الدراسة على تحسين من عملية اتخاذ القرار داخل المؤسسات والمساهمة في تحسين إستراتيجيات الموارد البشرية استنادا إلى نتائج نموذج التنبؤ بسلوك الموظفين.
- التنبؤ بسلوك الموظفين داخل المؤسسة يساهم في تطوير وتحسين رضاهم و ادائهم ويزيد من ولائهم للشركة.
- التنبؤ بسلوك الموظفين يساعد المؤسسات الاقتصادية على اتخاذ قرارات ناجحة و استباقية كما يقلل من تكاليفها و يقلل من المخاطر ويحقق لها الاداء مستدام.
- قدرة تقنيات تعلم الآلة في التعامل مع كم هائل من البيانات ومعالجتها واكتشاف الانماط في البيانات مثل العلاقات غير الواضحة بين المتغيرات.

6- أهداف الدراسة:

- دراسة المفاهيم الأساسية التي تساعد على فهم سلوك الموظفين داخل المؤسسة.
- التعرف على أهم خوارزميات تعلم الآلة التي يمكن استخدامها في بناء نموذج تنبؤ بسلوك الموظفين داخل المؤسسة.
- بناء نموذج تنبؤ بسلوك الموظفين بناءا على بيانات خاصة بشركة اي بي ام.
- تقييم أداء النموذج والتأكد من أنه يمكن الاعتماد عليه في التنبؤ من خلال معايير خاصة بقياس الدقة مثل معيار (Accuracy).
- مقارنة بين عدة نماذج لتعلم الآلة واختيار الأفضل بينها من حيث الدقة.

7- منهج الدراسة:

تم الاعتماد في هذه الدراسة على المنهج الوصفي من خلال وصف البيانات والتعرف على أهم المفاهيم والمصطلحات المتعلقة بالدراسة و المنهج التحليلي من خلال تحليل بيانات الموظفين و خصائصهم و ميزاتهم باستخدام تقنيات تعلم الآلة والتنبؤ بسلوكهم في المستقبل.

الفصل النظري

الفصل الأول : الإطار النظري

أصبح في العصر الحالي توجه كبير نحو دمج ادارة الأعمال و ادارة الموارد البشرية بالتكنولوجيا و أحدث التقنيات في تحليل و فهم سلوكيات الموظفين و الذي يعتبر هذا الأخير بالغ الأهمية بالنسبة للمؤسسات الاقتصادية بسبب أنه يؤثر على انتاجيتها و كفاءتها حيث يجب على المؤسسات الاقتصادية دراسة سلوك الموظفين و فهم اهم المحددات و العوامل التي تلعب دورا في التأثير على سلوك موظفيها و في هذا العصر توفرت أدوات و تقنيات حديثة مثل تعلم الآلة حيث أصبحت ضرورية و أساسية في فهم سلوك الموظفين بل حتى التنبؤ بهذه السلوكيات و كيف ستكون مستقبلا مما يكسب المؤسسة ميزة تنافسية و التفوق على منافسيها. وفي هذا الفصل سنتعرف الى مفهوم سلوك الموظفين و أهم محدداته و أهمية دراسته و السلوك الذي يتنبأ به. أيضا سنتطرق الى مدخل لتعلم الآلة و اهم نماذجه و تقنياته و مجالات استخداماته و لماذا هو ضروري في مجال ادارة الاعمال و ادارة الموارد البشرية خاصة فيما يخص سلوك الموظفين.

المبحث الاول: مدخل الى سلوك الموظف داخل المنظمة:

يعتبر سلوك الموظف داخل المنظمة عاملا أساسيا في تحقيق أهدافها، إذ يعكس مدى تفاعل الأفراد مع بيئة العمل والسياسات التنظيمية، فالمنظمة تعتمد على العنصر البشري بوصفه المحرك الرئيس لأنشطتها، مما يجعل فهم دوافع واتجاهات الموظفين ضرورياً لتحسين الأداء وتعزيز الرضا الوظيفي. كما أن دراسة هذا السلوك تساعد الإدارة على تطوير استراتيجيات فعالة لبناء ثقافة تنظيمية إيجابية تدعم النجاح والاستمرارية في مواجهة تحديات بيئة الأعمال.

المطلب الاول: مفهوم سلوك الموظف

الفرع الاول: تعريف سلوك الموظف

1- السلوك: السلوك هو ذلك التفاعل الذي يحصل بين الانسان و بيئته المحيطة و حيث يكون في الغالب سلوك يكتسبه الانسان من خلال الملاحظة و التعلم و التدرب حيث يتعلم الانسان من سلوكيات بسيطة و من سلوكيات معقدة جدا و حيث كلما كان السلوك الذي يفعله الانسان منضبطا و عملي و مقبول كان ذلك السلوك ايجابيا.[1]

[1] فاروق عبد الفتاح موسى ، **اسس السلوك الانساني**، دار عالم الكتاب، 1985، الصفحة 8

2- الموظف: هو ذلك الشخص الذي يتعين عليه القيام بوظيفة معينة و الذي تم ترسيمه في وظيفة و تدرج متسلسل سواءا كان ذلك في الادارات العمومية او في المؤسسات الاقتصادية و الشركات التي يعمل فيها الشخص.[1]

3- سلوك الموظف: سلوك الموظف هو تلك الاستجابة التي تصدر من جانب الموظفين اتجاه المواقف المعينة في بيئة العمل او المكان الذي يكون فيه الموظفين متواجدون فيه بصفة رسمية حيث يمكن للمهام و الاوامر المسندة للموظفين او الظروف المحيطة ببيئة العمل ان تثير سلوكيات مختلفة بشكل كامل بين الموظفين حيث يمكن ان تكون هذه السلوكيات ايجابية للمؤسسة او يمكن ان تكون سلبية للمؤسسة. [2]

الفرع الثاني: أهمية دراسة سلوك الموظف في المؤسسة

تتمثل في:[3]

- تعطي دراسة الموظفين رؤية و صورة واضحة حول سلوك الافراد و الموظفين داخل المؤسسات التي تختلف حسب الظروف حيث يمكن للمدير ان يأخذ بذلك فكرة و نظرة واضحة عن كيفية اداء موظفيه للمهام و هل يعودون بمردودية او لا على المؤسسة مما يمنحه نوعا من اليقضة.

- محاولة فهم لماذا يقوم أي عامل بسلوك معين وما هو السبب في ذلك بمعنى معرفة الدوافع التي تدفع الإنسان إلى سلوك معين وتعتبر عملية جوهرية في كيفية التعامل معه والسعي لتعديل سلوكه إذا كان فيه أي انحراف عن قواعد العمل الساري هبا العمل في المنظمة.

- التنبؤ بسلوك الموظفين داخل المؤسسة بحيث يستطيع المدير أن يصدر أحكاما دقيقة عن سلوك الموظفين في المؤسسة، وإسناد المسؤوليات والمهام إلى كل من يتوقع منه الأداء الجيد في موقع المسؤولية الذي يتواجد فيه.

- التحكم في سلوك الموظفين عن طريق التحفيز والعقاب إذا كان هناك انحراف أو تقصير في الأداء.

---

[1] بلعرابي عبد الكريم، **الحالات القانونية لانهاء خدمة الموظف العمومي في القانون الجزائري**، دفاتر السياسة والقانون، العدد 16، كلية الحقوق، جامعة ورقلة، جانفي 2017، ص196

[2] Inemesit Godwin Enyina,Samuel Okurebia and Kingsley Uwa,**Employee's Work Behaviour: Absenteeism, Embezzlement, & Organizational Performance**،International Journal of Business and Management Review2023,p86

[3] مولاي علي الزهر،مقال **السلوك التنظيمي بين النظرية والتطبيق**،مجلة الآداب والعلوم الاجتماعية،2017،العدد 17،ص4

– الغاية من دراسة سلوك الموظفين هو استعمال وسائل التأثير الايجابي في العامل، حتى يساهم بطريقة ايجابية في تحقيق الأهداف المؤسسة

المطلب الثاني: محددات سلوك الموظف

الفرع الاول: المحددات الفردية

1- العمر: هو الفترة الزمنية التي انقضت منذ ولادة الكائن الحي وحتى اللحظة الراهنة، ويُقاس عادةً بالسنوات أو الأشهر، لكنه لا يقتصر على البعد الزمني فقط بل يشمل أبعادًا بيولوجية، نفسية، واجتماعية. هذا المفهوم يُستخدم في علوم مثل علم الاجتماع، علم النفس، والبيولوجيا لفهم النمو، الشيخوخة، والمسؤوليات الاجتماعية المرتبطة بمراحل الحياة[1].

2- التوازن بين الحياة و العمل: التوازن بين العمل والحياة هو قدرة الموظفين على تحقيق التوازن بين الحياة الشخصية والعمل، وأن تظل فعالاً ومنتجاً في العمل، مع الحفاظ على حياة صحية وسعيدة في الجبهة الداخلية مع التمتع بأوقات الفراغ الكافية، على الرغم من وجود عمل الضغط والأنشطة التي لا نهاية لها والتي تتطلب الوقت والاهتمام .وبالتالي يمكن اعتبار التوازن بين العمل والحياة بمثابة التفاعل بين العمل مدفوع الأجر والعمل غير مدفوع الأجر داخل الأسرة والمجتمعات، فضلاً عن أوقات الفراغ والاسترخاء. و يعرف التوازن بين العمل والحياة ايضا بانه القدرة على الموظفين لتحقيق التوازن في مجالات العمل والأنشطة غير الوظيفية .[2]

3- الخبرة: هي نتاج تراكم معرفي ومهاري ونفسي واجتماعي، فهي من الناحية المعرفية تمثل القدرة على استخدام المعلومات والمعارف المكتسبة بفعالية في مواقف عملية، ومن الناحية المهارية تعكس المهارات التي يطورها الفرد عبر التدريب والممارسة المستمرة مما يميّز الخبير عن المبتدئ. أما من الناحية النفسية فهي ترتبط بالقدرة على اتخاذ قرارات دقيقة وسريعة في المواقف المعقدة نتيجة تراكم التجارب السابقة، ومن الناحية الاجتماعية والمهنية تُعد معيارًا للتميز والاعتراف بالكفاءة داخل المجتمع ومجالات العمل والتعليم. وبذلك تُفهم الخبرة على أنها إطار شامل يجمع بين المعرفة والمهارة والتجربة والاعتراف الاجتماعي، مما يجعلها أساسًا للتفوق والإبداع في مختلف المجالات[3].

[1] السويدان، عبد الرحمن. علم النفس النمو: الطفولة والمراهقة. دار الفكر، 2015، ص. 22.

[2] Mazin Aledeinat, Zeyad Alkhazali, Hakem Sharari, Malek Alsoud**, TOWARDS A NEW MODEL FOR WORK-LIFE BALANCE: A FACTORS-REVIEW APPROACH**, Journal of Management Information and Decision Sciences,2021,p2

[3] الزيات، أحمد. علم النفس التربوي وتطبيقاته في التعليم. القاهرة: دار الفكر العربي، 2018.

4- الرضا الوظيفي: الشعور المتولد لدى الفرد نتيجة استجابته للموقف الكلي للعمل اي ن الرضا الوظيفي هو "الاتجاه العام نحو وظيفة معينة" أو هو الاختلاف بين مقدار العوائد المستلمة من قبل العاملين وذلك المقدار الذي يفترض أن يستلمه[1].

الفرع الثاني: المحددات التنظيمية

تتمثل المحددات التنظيمية في الآتي:[2]

1- الرواتب: يُقصد بالراتب الأموال التي يحصل عليها الشخص ويستخدمها لإعالة نفسه وتلبية احتياجاته. ويُعتبر الراتب دافعًا لتحقيق النجاح حيث يُمنح الراتب للموظفين لكي يتمكنوا من العمل بأفضل ما لديهم والمساهمة في تحسين أداء الشركة. لذلك، ومن أجل تحفيز الموظفين، تحتاج الشركات إلى إيلاء اهتمام كبير برواتب العاملين فإذا تجاهلت الشركات مسألة الرواتب، فقد تظهر العديد من المشكلات. ومن أخطر هذه المشكلات احتمال انتقال الموظفين ذوي الأداء المتميز إلى شركات أخرى.[3]

2- الترقيات: هي انتقال الموظف من وظيفة إلى أخرى أعلى منها في السلم الوظيفي، سواء من حيث المسؤوليات أو الصلاحيات أو الأجر، وتُعد وسيلة أساسية لتحفيز العاملين ورفع مستوى أدائهم. وهي لا تقتصر على البعد الإداري فقط، بل تشمل أيضًا أبعادًا اجتماعية (تحقيق المكانة والاعتراف الاجتماعي)، ونفسية (تعزيز الرضا الوظيفي والدافعية)، واقتصادية (زيادة الدخل وتحسين الوضع المعيشي). وبذلك تُفهم الترقية على أنها عملية إدارية واجتماعية ونفسية متكاملة تهدف إلى تطوير الفرد والمؤسسة في آن واحد.

3- بيئة العمل: هي مجموعة الظروف والعوامل المادية والتنظيمية والاجتماعية والنفسية التي تحيط بالعامل داخل المؤسسة وتؤثر في أدائه وإنتاجيته ورضاه الوظيفي. وتشمل هذه البيئة عناصر مثل المكان المادي (المكاتب، الإضاءة، التهوية)، والعلاقات الاجتماعية بين الزملاء والإدارة، إضافةً إلى السياسات التنظيمية والثقافة المؤسسية. ومن الناحية النفسية، تُعد بيئة العمل عاملًا مهمًا في تعزيز الدافعية، الانتماء، والإبداع، بينما من الناحية الاجتماعية فهي تساهم في بناء شبكة من العلاقات المهنية التي تدعم التعاون والانسجام داخل المؤسسة.

---

[1] ME,Sermpene et Al,Job Satisfaction in Relation to Organizational Cultur,Journal of Industrial Psychology, 28(2), 2002, p23.

[2] الشرقاوي، عبد الرحمن، **إدارة الموارد البشرية: مدخل استراتيجي**، القاهرة: دار الفكر العربي، 2019. ص87.

[3] Nugraha, Putri, Mubarok, Hadani, **The Effect of Salary on Employee's Performance: A Study in Central Java**, Proceedings of the 7th Global Conference on Business, Management, and Entrepreneurship (GCBME 2022),2024,p1812,1814

4- القيادة: هي عملية التأثير في الأفراد وتوجيههم نحو تحقيق أهداف مشتركة، من خلال القدرة على تحفيزهم، تنظيم جهودهم، واتخاذ القرارات المناسبة. وهي لا تقتصر على البعد الإداري فقط، بل تشمل أيضًا أبعادًا اجتماعية (بناء العلاقات والثقة بين القائد وأتباعه)، ونفسية (إلهام الأفراد وتعزيز دافعيتهم)، وتنظيمية (توزيع المهام وتنسيق الموارد). وبذلك تُفهم القيادة على أنها فن وعلم يجمع بين التأثير الشخصي والمهارة الإدارية لتحقيق النجاح الفردي والجماعي.

المطلب الثالث: سلوكيات الموظفين القابلة للتنبؤ

1- الغياب الوظيفي: هو غياب العامل عن عمله لمدة يوم او اكثر سواء كان بإذن او بدون و لا تحسب ضمن غياب الايام التي يعمل فيها الفرد بسبب توقف الانتاجية او الاجازات السنوية، وقد يكون فعل لا ارادي تسببه قوة قاهرة ناتجة عن اسباب مرضية او غيرها من الاسباب الاخرى[1].

2- الاداء الوظيفي: لقد برزت عدت تعريفات للأداء الوظيفي نذكر منها ما يلي[2]:

- يشير الاداء الوظيفي الى درجة تحقيق واتمام المهام المكونة لوظيفة الفرد، وهو يعكس الكيفية التي يتحقق بها، او يشبع الفرد بها متطلبات الوظيفة وغالبا ما يحدث لبس و تداخل بين الاداء و الجهد، فالجهد يشير الى الطاقة المبذولة، اما الاداء فيقاس على اساس النتائج التي يحققها الفرد.
- هو الاثر الصافي لجهود الفرد التي تبدأ بالقدرات وإدراك الدور أو المهام و بالتالي يشير إلى تحقيق وإنجاز المهام المكونة لوظيفة الفرد.
- يعرف ايضا بانه السلوك الذي يقوم به المعلم في المؤسسة التي يعمل فيها بالطريقة الصحيحة والسليمة، مراعيًا بذلك الفعالية والكفاءة والسلامة العامة في العمل.

3- الالتزام التنظيمي: لقد برزت عدت تعريفات للالتزام التنظيمي نذكر منها ما يلي[3]:

- هو الرغبة الداخلية للفرد في المنظمة لبذل أقصى درجات الجهد في العمل والبالغ من الإيمان بأهداف وقيم المنظمة والرغبة الشديدة بالمحافظة على عضويتها وذلك بهدف الحفاظ على استمراريتها وتطويرها"، والالتزام التنظيمي لديه خاصية مركبة من أربع مكونات أساسية وهي: الولاء التنظيمي، المسؤولية اتجاه المنظمة، الرغبة بالاستمرار بالعمل في المنظمة والإيمان بالمنظمة.

---

[1] عادل حسن، **الكفاية الانتاجية للأفراد في الصناعة**، دار الجامعات المصرية، الإسكندرية، 1969، ص202.

[2] اسعد احمد محمد عكاشة، **اثر الثقافة التنظيمية على مستوى الاداء الوظيفي**، دراسة تطبيقية على شركة الاتصالات بفلسطين، رسالة ماجستير في ادارة الاعمال، غزة، 2008، ص33.

[3] رشيد خطارة، **الذكاء الانفعالي وعلاقته بالالتزام التنظيمي**، اطروحة دكتوراه علم النفس التربوي، ورقلة، الجزائر، 2019، ص60.

– وأشار كل من موداي، ستيرز و بورتر أن الالتزام التنظيمي يمثل "اعتقاداً قوياً من جانب الأفراد لأهداف المنظمة وقيمها ورغبتهم في بذل أكبر عطاء أو جهد ممكن لمصالح المنظمة التي يعملون بها مع رغبة قوية في الاستمرار في عضويتها.

– كما عرف كل من عبد الوهاب وخطاب، الالتزام التنظيمي بأنه استثمار متبادل بين الفرد والمنظمة باستمرار العلاقة التعاقدية، يترتب عليه سلوك الفرد سلوكاً يفوق السلوك الرسمي المتوقع منه والمرغوب فيه من جانب المنظمة، ورغبة الفرد في إعطاء جزء من أجل الإسهام في نجاح واستمرارية المنظمة، كالاستعداد لبذل مجهود أكبر، والقيام بأعمال تطوعية، وتحمل مسؤوليات إضافية.

4- خروج الموظفين: إنهاء الأعضاء لعلاقتهم بالمنظمات بشكل طوعي، و ايضا يعرف بانه نسبة العاملين الذين يغادرون المنظمة خلال فترة زمنية معينة قبل الموعد المتوقع لانتهاء عقودهم، و منع يعتبر تسرّب الموظفين هو معدل التغير في عدد العاملين داخل المؤسسة خلال فترة زمنية محددة. ويمكن أن يكون تسرّب الموظفين في أي منظمة إما طوعيًا أو غير طوعي، و هناك العديد من العوامل التي تؤثر في معدل تسرّب الموظفين في أي منظمة، وقد تكون هذه العوامل ناتجة عن صاحب العمل أو الموظفين أنفسهم. وتشمل هذه العوامل الأجور، ومزايا الشركة، ومواظبة الموظفين، والأداء الوظيفي، وجميعها تلعب دورًا مهمًا في تسرّب الموظفين.[1]

المبحث الثاني: عموميات حول التنبؤ و تقنيات تعلم الآلة

بعد ان تطرقنا في المبحث السابق حول سلوك الموظفين و اهم عوامله و اهم السلوكيات التي يمكن التنبؤ بها سنتطرق في هذا المبحث الى مفهوم التنبؤ و اهميته و مراحله و اهم طرقه من اجل فهم اختيارنا للتقنية التي يمكننا من خلالها التطبيق في هذه الدراسة و التي تعرف بالتعلم الآلي و سنتعرف الى اهم انواع هذه التقنية و اهم استخداماتها و ايضا علاقتها بإدارة الموارد البشرية التي يندرج تحتها سلوك الموظفين.

[1] Eteng, Joy Umuolo،Bernard Samuel Eventus،Chenui, Armand Phobuh،Bassey Bassey Ubi،Okon, Utibe Essien & Etim،Umo Ephraim،**Employee Attrition and Organisational Sustainability**،International Journal of Social Sciences and Management Research،2025،p337

المطلب الاول: عموميات حول التنبؤ

أولا: تعريف التنبؤ

– التنبؤ يعرف على انه معرفة ماذا سيحدث في المستقبل بأكبر قدر ممكن بحيث تكون المعرفة حول ذلك دقيقة و موثوقة من خلال جمع المعلومات المتاحة و التي تتمثل في البيانات التاريخية و تحليلها و معرفة الاحداث المستقبلية.[1]

–التنبؤ هو تلك الرؤية او النظرة المستقبلية لما يحدث في المستقبل بناءا على ما يحدث على ارض الواقع و البيانات التاريخية و بيانات الحاضر مع معرفة الاسباب هو نظرة مستقبلية لما ستكون عليه الظواهر بناءا على الواقع أو بيانات الماضي والحاضر مع معرفة المتغيرات و العوامل التي تسبب الظاهرة المدروسة و الذي على اساس هذا ترسم المؤسسة انشطتها و خططها.[2]

ثانيا: أهمية التنبؤ: هناك بعض الاهميات للتنبؤ نذكر منها:[3]

– يعتبر التنبؤ اداة جوهرية و اساسية في الادارات الاستراتيجية في اي مجال كان حيث يشمل وضع خطط و استراتيجيات استعدادا لما سيحدث في المستقبل.

– يعمل التنبؤ على جعل المؤسسة تستجيب للاحتياجات المعلوماتية للمدراء الاستراتيجيين بما يتعلق بالبيئة و الظروف الخارجية.

– يتيح التنبؤ للمؤسسة الفرص لمختلف اداراتها لفهم ماذا يدور و ماذا يتحرك في البيئة المحيطة و الذي يعد هذا امرا حاسما لاتخاذ قرارات استراتيجية.[4]

– يعمل التنبؤ على شيء مهم جدا في الادارات الحديثة و هو تقليص حالة عدم اليقين و الشك حول ماذا سيحدث في المستقبل حيث يمكننا بذلك العمل على تخفيض و تقليص الانحراف المعياري لما تتوقعه المؤسسة في المستقبل.

---

[1] Rob J. Hyndman and George Athanasopoulos, **Forecasting: Principles and Practice**, 2014, p12

[2] بوغازي فريدة، **فعالية تطبيق تقنيات التنبؤ بالمبيعات في المؤسسة دراسة تطبيقية بمؤسسة GNL / سكيكدة بالجزائر**،ص76،مجلة الباحث الاقتصادي،العدد 04 / ديسمبر 2015

[3] Cihat Polat, **Forecasting as a Strategic Decision Making Tool: A Review and Discussion with Emphasis on Marketing Management**, European Journal of Scientific Research, 2008, p. 348

[4] Régis Bourbonnais and Jean-Claude Usunier, **Sales Forecasting: Between Theory and Practice**, 2008, p36

– يعمل التنبؤ على المساهمة في معرفة امكانية حدوث التقلبات الاقتصادية بشكل عام و انتقال مرحلة النمو الى مرحلة الركود او العكس او اذا كان سيحدث تقلب اقتصادي حديث بحيث هل سيحدث ذلك او لا.

ثالثا: مراحل التنبؤ

نذكر منها مايلي:[1]

المرحلة الاولى: فهم المشكلة وتعريفها

تبدا عملية التنبؤ من خلال تحديد السؤال المشكلة المطروحة بدقة لان هذه المرحلة تعتبر اساسية في بناء حل مناسب للمشكلة المطروحة حيث يتطلب ذلك الفهم الكبير للإمكانيات المتوفرة و المتطلبات و النتائج المرجوة وعند غموض المتطلبات، ينبغي تحديد المتغير المراد التنبؤ به والتأكد من صلته بالمشكلة.

المرحلة الثانية: جمع البيانات

تعتبر مرحلة جمع البيانات احد اكثر المراحل صعوبة و تعقيدا في مراحل التنبؤ حيث يعتمد جمع البيانات على استرداد البيانات من مصادر كثيرة و متنوعة مع ضرورة توفر بيانات اضافية و جديدة من اجل ضمان الحل الامثل للمشكلة المقترحة و ايضا يجب ان تكون هذه البيانات ملائمة و حديثة و تتوافق مع الهدف الذي يجب تحقيقه.

المرحلة الثالثة: معالجة البيانات

في اغلب الاحوال تكون البيانات الخام التي تم جمعها من مصادر مختلفة غير جاهزة و نظيفة للتحليل و لذلك تعتبر مرحلة تنظيف و معالجة البيانات مرحلة ضرورية و اساسية حيث يقضي المحللون و علماء البيانات وقتا طويلا في معالجة و تنظيف البيانات من اجل تصحيح اخطائها و حذف القيم المكررة و الشاذة فيها و كل هذا من اجل ان تكون البيانات سليم و جاهزة للتحليل اذ تعتمد جودة التنبؤ على جودة البيانات التي تم تمريرها للتنبؤ حيث يساعد هذا على رفع دقة التنبؤ و زيادة درجة موثوقيته.[2]

[1] Khirod Chandra Panda and Shobhit Agrawal, **Predictive Analytics: An Overview of Evolving Trends and Methodologies**, Journal of Scientific and Engineering Research, 2021, p178

[2] Venkata Kalyan Chakravarthy Mandavilli, **Harnessing the Power of Predictive Analytics: Transforming Business Intelligence**, European Journal of Computer Science and Information Technology, 2025, p2

المرحلة الرابعة: تحليل البيانات الاستكشافي

يعتبر التحليل الاستكشافي للبيانات مرحلة مهمة ايضا حيث نمر من خلالها من اجل ان نفهم خصائص البيانات و اكتشاف الانماط المخفية و اكتشاف كيف تتوزع البيانات و اكتشاف الارتباطات ايضا و تنقسم هذه المرحلة الى اربعة انواع اساسية و هي التحليل الاحادي حيث نصف متغيرا واحدا بدون الرسومات احصائيا و التحليل الاحادي بالرسومات البيانية مثل المدرجات التكرارية والمخططات الصندوقية، ومتعدد المتغيرات بدون الرسومات البيانية الذي يحلل العلاقات بين عدة متغيرات عبر الجداول الإحصائية، وأخيراً متعدد المتغيرات بالرسومات البيانية الذي يبرز الروابط المعقدة بين البيانات بصرياً كالمخططات الشريطية المجمعة.

المرحلة الخامسة: بناء النموذج

تأتي مرحلة بناء النموذج بعد ان تم تحليل البيانات و استكشافها باستخدام طرق و نماذج التنبؤ حيث في هذه المرحلة بالتحديد نختار المتغير التابع و الذي سنتنبؤ به و نختار ايضا المتغيرات المستقلة التي تستخدم في التنبؤ بالمتغير من خلالها. و نستخدم عدة طرق للتنبؤ من اجل الوصول افضل النتائج المطلوبة.

المرحلة السادسة: تقييم النموذج

تأتي هذه الخطوة بعد ان قمنا ببناء النموذج حيث تتمثل هذه الخطوة في فحص و التأكد كم مدى فعالية النموذج الذي قمنا ببنائه حيث يساعد تقييم النموذج على اختيار النموذج المثل و الافضل من بين النماذج لمعالجة المشكلة المطروحة .ويتم تقييمه عبر عدة مقاييس للتقييم.

رابعا: طرق التنبؤ

1- الطرق النوعية:

يتم استخدام الطرق النوعية عندما لا تتوفر لدينا بيانات تاريخية مثل في حالة تقديم منتج جديد لا تتوفر لديه بيانات تاريخية، وتعتمد هذه الطرق النوعية على استثمار الحكمة و التجربة التي تمتلكها

الإدارة، فضلا عن مجموعة من العوامل الأخرى والمعلومات التي يمتلكها الأفراد كالحدس والخبرة الشخصية، و علة هذا النحو سنذكر بعض منها كمايلي: [1]

ا- طريقة دلفي:

تعتبر طريقة دلفي بانها ذلك الاسلوب الذي يعتمد على الوصول إلى توافق بين آراء مجموعة من الخبراء حول التنبؤ بحدوث ظاهرة أو حدث في المستقبل، مع الحفاظ على سرية هوية المشاركين. أي انها اسلوب يقدم فيه العضو او الخبير رايه حول ظاهرة معينة دون ان يعرف هوية باقي الاعضاء، و يتم اختيار الخبراء الذي يقدمون آرائهم بسرية تامة و من بلدان مختلفة لتجنب الانحياز .

ب- استطلاع آراء المستهلكين:

تتمثل طريقة استطلاع اراء المستهلكين على جمع تصورات و اراء المستهلكين حول ترتكز هذه حول الطلب المتوقع خلال فترة محددة، عادة بين شهر وسنة، و يتم ذلك من خلال وسائل مثل المقابلات المباشرة، أو الاستبيانات، أو لقاءات مع العملاء.

ج- استطلاع رأي الإدارة العليا:

تعتبر طريقة استطلاع اراء الادارة العليا اسلوب يقوم أعضاء الإدارة العليا أو مجموعة منهم بتقديم تقديراتهم حول مستوى الطلب المتوقع خلال الفترة الزمنية القادمة، بعدها يتم تقدير هذه التوفيقات بحساب المتوسط الحسابي لها لكي يعتمد هذا المتوسط الحسابي كمؤشر لحجم الطلب او القيمة متوقعة مستقبلا .

د- أبحاث السوق:

تعتبر طريقة ابحاث السوق عملية تهدف الى وضع فرضيات تتعلق بالسوق و اختبارها، كما تعد هذه الطريقة اداة مهمة تساعد إدارة العمليات على جمع معلومات حول نوايا و ميول المستهلكين الشرائية في المستقبل. يمكن لطريقة ابحاث السوق ان توفر بيانات اساسية تساهم في دهم عملية التخطيط و تطوير المنتجات بدلا من جمع اراء و ميول المستهلكين فقط.

[1] حنان بن عوالي،مقال **التنبؤ بالطلب كمؤثر أساسي في عملية تخطيط الإنتاج**، مجلة الباحث الإقتصادي،ص206،207، العدد 01،2013

2- الطرق الكمية:

تعتمد الطرق الكمية للتنبؤ على البيانات من اجل التنبؤ بدلا من التقدير البشري الحدسي. يتم اختيار النموذج الكمي للتنبؤ بناء على عدة عوامل مختلفة مثل سياق عملية التنبؤ ومدى وفرة البيانات التاريخية وعلاقتها ارتباطها بموضوع التنبؤ، ومستوى الدقة المطلوب، إضافة إلى الفترة الزمنية التي يغطيها التنبؤ.[1]

و تتمثل الطرق الكمية فيما يلي:

ا- نماذج السلاسل الزمنية

تعتبر السلسلة الزمنية مجموعة من الملاحظات التي تم التقاطها و التي تكون خاصة بظاهرة معينة تم تسجيلها في اوقات محددة و معينة و في الغاب تكون هذه الفترات المحددة متساوية بظاهرة معينة يتم تسجيلها في أوقات محددة، وغالباً ما تكون هذه الفترات متساوية. يتم استخدام السلسلة الزمنية في تحليل مختلف التغيرات التي تحدث في ظاهرة معينة من اجل اكتشاف و تحديد انماطها و اسبابها المختلفة ومعرفة اتجاهاتها، و هذا يجعلنا نستعد للأحداث المستقبلية بشكل افضل و ادق و هذا ما يجعل السلسلة الزمنية مهمة.[2]

ب- نماذج التعلم الآلي:

وهي الطريقة الاحدث و الافضل في العصر الحالي حيث تعتبر الطريقة المثالية لتحليل مسائل التنبؤ المعقدة كالتصنيف الثنائي و التنبؤ متعدد المتغيرات، وهي التي سنستخدمها في هذه الدراسة.

المطلب الثاني: مفاهيم تعلم الآلة

من خلال المطلب السابق علمنا ان تعلم الآلة يعتبر احد الطرق المهمة من طرق التنبؤ و ليس ذلك فقط بل يعتبر اكثر الادوات تطورا و كفاءة في ذلك حيث اصبح الوقت الحالي اهم اداة في التنبؤ حيث سنتطرق لأنواعه و خوارزمياته.

[1] Maximilian Zellner, Ali Abbas, David Budescu, and Aram Galstyan**, A survey of human judgement and quantitative forecasting methods**, Royal Society Open Science, 2021, p11

[2] حنان بن عوالي،نفس المرجع السابق،ص210

اولا: تعريف تعلم الآلة

– التعلم الآلي هو فرع من فروع الذكاء الاصطناعي حيث يهدف الى فهم بنية البيانات وملائمة البيانات في النماذج التي يمكن فهمها واستخدامها من قبل المتخصصين و الاشخاص.[1]

– يعتبر تعلم الالة مهارة اساسية و قوية يمكننا من خلاله السماح للآلات بالتعلم بشكل مستقل من خلال اعطائها البيانات الخاصة بالجهاز، و مع التطور التكنولوجي الكبير في العصر الحالي يتم تدريب الآلات و جعلها تتصرف مثل البشر في القدرة على اتخاذ القرار.[2]

ثانيا: انواع التعلم الآلي

1- التعلم الخاضع للإشراف:

يعمل التعلم الخاضع للإشراف على جعل النموذج يتعلم من بيانات تدريب معلمة ،حيث يتيح لنا هذا الاخير اجراء و عمل تنبؤات حول بيانات مجهولة او غير مرئية أو مستقبلية. ويشير مصطلح "خاضع للإشراف" إلى مجموعة من مدخلات البيانات التي يكون اسم المخرجات المطلوبة (التسميات) فيها معروفة مسبقا.[3]

1-1- فئات التعلم الخاضع للإشراف:

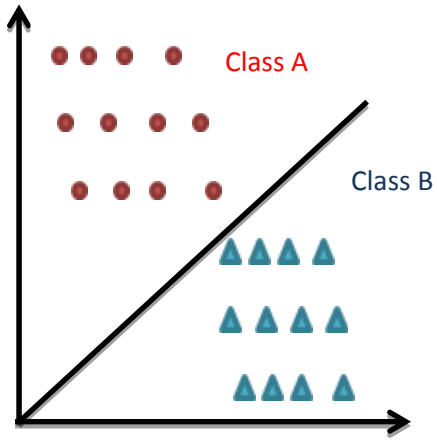
ا- التصنيف:

التصنيف (Classification) هو فئة فرعية من التعلم الخاضع للإشراف، حيث يكون الهدف هو التنبؤ بالتسميات الفئوية للعينات الجديدة اعتمادًا على الملاحظات السابقة.[4]

وعموما يستخدم التصنيف من أجل :تصفية البريد الالكتروني من الرسائل المزعجة و كشف عن اللغة المستخدمة والبحث عن وثائق مماثلة وتحليل المشاعر و الكشف عن الغش.[5]

[1] ليزا تاغليفيري، كتاب **مشاريع تعلم الآلة بايثون**،2022،ص15
[2] ميلاد وزان ،**كتاب تعلم الالة و علم البيانات** ،ص36
[3] Sebastian Raschka and Vahid Mirjalili, **Python Machine Learning**, 2019, p3
[4] Sebastian Raschka and Vahid Mirjalili ،نفس المرجع السابق،ص3
[5] محمد لحلح، كتاب **مدخل إلى الذكاء الاصطناعي وتعلم الآلة**،2020، ص69

الشكل: شكل التصنيف



المصدر: من اعداد الطالب

ب- الانحدار:

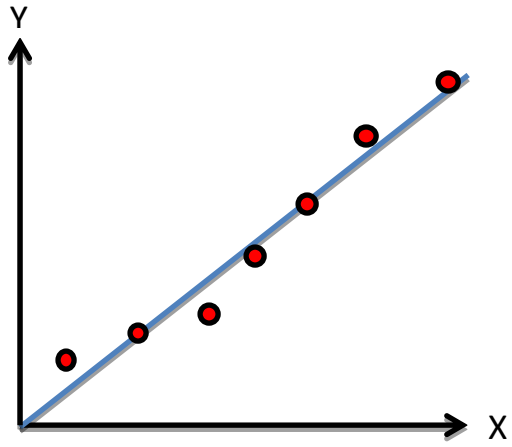
- النوع الثاني من التعلّم الخاضع للإشراف فهو التنبؤ بالمخرجات المستمرة، والذي يُعرف أيضًا بـ "تحليل الانحدار" حيث في هذا الاخير يعطى لنا عدد من المتغيرات التفسيرية أي المتغيرات المستقلة ومتغير الذي نتنبؤ به أي المتغير التابع، ونحاول إيجاد علاقة بين هذه المتغيرات من اجل التنبؤ بالنتيجة المستقبلية[1].

- يعتبر الانحدار ايضا اساس آلية للتصنيف ولكن هنا نتوقع رقما بدلاً من فئة. ومن الامثلة على ذلك توقع سعر السيارة من خلال المسافة المقطوعة، وتوقع حركة المرور بحسب وقت محدد من اليوم، وتوقع حجم الطلب من خلال نمو الشركة و توقعات أسعار الاسهم و تحليل حجم الطلب والمبيعات[2].

[1] Sebastian Raschka and Vahid Mirjalili, **Python Machine Learning**, 2019, p4

[2] محمد لحلح، نفس المرجع السابق، ص74

الشكل: شكل الانحدار



المصدر: من اعداد الطالب

1-2- خوارزميات التعلم الخاضع للإشراف:

1- الانحدار اللوجستي:

الانحدار اللوجستي هو أحد النماذج الخطية المعمّمة، ويُستخدم عندما تكون القيم الناتجة غير مستمرة. يُطبَّق غالبًا لتقدير احتمال انتماء البيانات إلى فئات مختلفة ضمن ما يُعرف بالتصنيف، سواء كان ثنائيًا (نعم/لا) أو متعددًا أو ترتيبيًا (ضعيف/متوسط/جيد).[1]

امثلة حول استخدامات الانحدار اللوجستي :[2]

- يُستخدم الانحدار اللوجستي بصفة عامة في الدراسات التي تهدف إلى التنبؤ وشرح قيم متغيرات كيفية، حيث يُعد من أنسب الأساليب الإحصائية في النماذج التنبؤية التي يكون فيها المتغير التابع نوعيًا.

- كما يُوظَّف هذا الأسلوب في التنبؤ باحتمال حدوث حدث معيّن قد يكون مرتبطًا بمخاطر، لذلك يُستعمل بكثرة في المجال الصحي والطبي للتنبؤ بإمكانية الإصابة بالأمراض وتحديد عوامل الخطر وعوامل القابلية للإصابة.

- يُستخدم الانحدار اللوجستي في مجالات متعددة من الحياة، خاصة في الطب والعلوم الاجتماعية، كما يُطبق في مجال التسويق لتقدير احتمال إقبال المستهلك على شراء منتج معين أو عزوفه عن الشراء.

[1] Craig Starbuck, **The Fundamentals of People Analytics**, Springer, 2023, p223

[2] محمد أمين دعيش، محمد ساري،مقال **نموذج الانحدار اللوجستي: مفهومه، خصائصه، تطبيقاته مع مثال تطبيقي الانحدار اللوجستي ثنائي الحدين على برنامج spss**،جملة السراج في التربية وقضايا المجتمع،العدد الاول،2017،ص125

2- شجرة القرار:

تُعد شجرة القرار إحدى طرق التعلم الخاضع للإشراف التي يمكن استخدامها لحل مشكلات التصنيف و الانحدار، حيث يعتمد هيكلها التنظيمي على شكل شجري يتكون من:

- العقد الداخلية : تعبّر عن السمات أو الخصائص الأساسية للبيانات.

- الفروع : تجسد المسارات التي تُتخذ بناءً عليها القرارات.

- عقد الأوراق : هي المحطة الأخيرة التي تمنحنا النتيجة النهائية للتصنيف.

تعمل هذه الشجرة كتمثيل مرئي وشامل لكافة السيناريوهات الممكنة استجابةً لموقف أو مشكلة معينة وفقاً للبيانات المتاحة. وبصورة مبسطة، تعتمد آلية عملها على طرح أسئلة محددة، ليتم لاحقاً تقسيم الشجرة إلى مسارات فرعية بناءً على الإجابة، والتي غالباً ما تكون بصيغة (نعم/لا).[1]

امثلة حول استخدامات شجرة القرار: [2]

- يمثل قطاع الرعاية الصحية أحد الميادين الرائدة في توظيف أشجار القرار، حيث اعتمد عليها الباحثون كأداة حيوية للتنبؤ بتشخيص الأمراض، وتقدير نتائج البروتوكولات العلاجية، وتحديد الإنذار الطبي لحالات المرضى. تكمن القوة الجوهرية لهذه الأشجار في كفاءتها العالية على استخلاص الأنماط والارتباطات المعقدة الكامنة في البيانات الطبية، الأمر الذي يساهم في الوصول إلى تشخيصات أكثر دقة ووضع خطط علاجية مصممة خصيصاً لتناسب احتياجات كل مريض.

- حققت أشجار القرار انتشاراً واسعاً في القطاع المالي؛ فمن خلال معالجة البيانات التاريخية واستنباط المتغيرات المؤثرة، تمنح هذه الخوارزميات قدرة دقيقة على التنبؤ بـ الجدارة الائتمانية للأفراد. وتشكل هذه المخرجات مرجعاً جوهرياً للبنوك وجهات الإقراض لتقييم المخاطر المحتملة قبل الموافقة على التسهيلات الائتمانية. وإلى جانب ذلك، تلعب أشجار القرار دوراً محورياً في منظومة مكافحة الاحتيال، حيث تقوم بفحص سجلات المعاملات المالية لرصد الأنماط المريبة، مما يساهم بشكل مباشر في الحد من الخسائر المالية وتأمين العمليات.

[1] Sonal Vinayak Kasari and Yash Arun Mali, **A Comparative Analysis of Logistic Regression and Decision Tree Algorithms for Fake News Detection**, International Journal of Novel Research and Development, 2023, p684

[2] Ibomoiye Domor Mienie and Norbert Jere, **A Survey of Decision Trees: Concepts, Algorithms, and Applications**, IEEE Access, 2024, p 8, 10

3-الغابة العشوائية:

تُعدّ الغابة العشوائية تطويرًا لشجرة القرار. وكما يدل اسمها، فهي تتكون من عدد كبير من أشجار القرار التي تعمل معًا ضمن نموذج واحد. تُسمّى عملية تدريب عدة نماذج على أجزاء مختلفة من بيانات التدريب ثم دمج نتائجها بالتجميع (bagging). ويعتمد هذا الأسلوب على فكرة "حكمة الجماعة"، حيث إن وجود عدد كبير من الأشجار غير المرتبطة التي تعمل بشكل جماعي يُفترض أن يعطي أداءً أفضل من الاعتماد على شجرة واحدة فقط.[1]

امثلة حول استخدامات الغابة العشوائية:[2]

– تُوظف هذه الخوارزمية لتقييم الجدارة الائتمانية للمتقدمين بطلبات القروض، مما يدعم المؤسسات المالية في اتخاذ قرارات مدروسة حول منح القرض أو رفضه. بالإضافة إلى ذلك، تُستخدم الغابات العشوائية بفاعلية في كشف ومكافحة الأنشطة الاحتيالية.

– يعتمد المتخصصون في المجال الطبي على أنظمة الغابات العشوائية لتشخيص الحالات المرضية من خلال تحليل السجلات الطبية السابقة للمرضى، كما تُسهم مراجعة هذه السجلات في تحديد الجرعات الدوائية المناسبة والدقيقة.

– يستخدم المحللون الماليون الخوارزمية لتحديد الأسواق المحتملة لتداول الأسهم، فضلاً عن دورها في مراقبة وتحليل سلوك واتجاهات الأسهم في السوق.

– تمكن خوارزميات الغابات العشوائية البائعين في منصات التجارة الإلكترونية من التنبؤ بتفضيلات المستهلكين، وذلك بالاعتماد على تحليل سلوكهم الشرائي واستهلاكهم السابق.

4- التدرج المعزز:

تعتبر خوارزميات التدرج المعزز (GBM) من الأساليب الفعّالة في التعلم الخاضع للإشراف ضمن مجال التعلم الآلي، إذ تُوظّف أشجار القرار بصفتها نماذج أولية (متعلمين ضعفاء) في سياق آلية التعزيز. ويسعى هذا النموذج إلى صياغة نظام تنبؤي عالي الدقة عبر معالجة أخطاء الأشجار

[1] Craig Starbuck, **The Fundamentals of People Analytics**, Springer, 2023, p246
[2] Hassan Ahmed Salman, Ali Kalakesh, and Amani Steiti, **Random Forest Algorithm Overview**, Babylonian Journal of Machine Learning, 2024, p74

السابقة بأسلوب تتابعي، بحيث تُصمم كل شجرة إضافية خصيصاً لسد الثغرات وتقليل البواقي التي نتجت عما سبقها من أشجار.[1]

امثلة حول استخدام التدرج المعزز:

– مسابقات الوصول لأعلى دقة تنبؤية في موقع kaggle و يعتبر هذا من اهم الاستخدامات على الاطلاق لهذه الخوارزمية.

– تستخدم في كثير من المجالات التي تستخدم فيها اشجار القرار و الغابة العشوائية.

5– آلة المتجهات الداعمة:

تعد خوارزمية الدعم الشعاعي من خوارزميات التعلم الخاضع للإشراف حيث تعتبر من اشهر الخوارزميات المستعملة في التصنيف الآلي وتعتمد على إنشاء إطار خطي أو غير خطي يقوم هذا الخط بالفصل بين فئات البيانات .[2]

امثلة حول استخدامات آلة المتجهات الداعمة: [3]

– تُستخدم خوارزمية آلة المتجهات الداعمة في المجال الطبي لتقديم تشخيص سريع وموثوق للأمراض. ويُعد الكشف المبكر عن المرض أمرًا مهمًا، إذ ان اكتشاف العدوى في وقت مبكر يساعد على إنقاذ العديد من الأرواح.

– يتم توظيف خوارزمية آلة المتجهات الداعمة في العديد من الدراسات الخاصة بتحليل المشاعر، حيث تتيح تصنيف الآراء إلى إيجابية وسلبية ومحايدة. كما تُستعمل نتائج هذا التحليل في مجالات مثل السياسة والاقتصاد والجريمة والعلاقات الدولية.

[1] Yue Liu a، Shulei Wu a، Zhongqiang Wu ، Shuangshuang Zhou ، **Application of gradient boosting machine in satellite-derived bathymetry using Sentinel-2 data for accurate water depth estimation in coastal environments** ، Elsevier،2024،p5

[2] فضيلة فراحي،مقال **استخلاص المتغيرات الاقتصادية المحفزة لنشاط بورصة الجزائر باستخدام خوارزمية متجه الدعم الآلي**،مجلة Les Cahiers du Cread،2022،المجلد 38،ص159

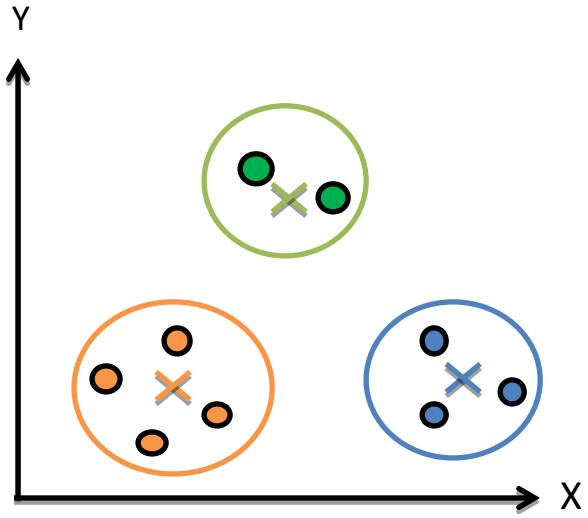
[3] Dkhaz Mustafa Abdullah, **Machine Learning Applications based on SVM Classification: A Review**, Qubahan Academic Journal, 2021, p82

2- التعلم الغير خاضع للإشراف:

في أسلوب التعلم غير الخاضع للإشراف، لا نزود الخوارزمية بالإجابات الصحيحة، أي أن البيانات لا تكون مُصنّفة. وبدلًا من ذلك، تعمل الخوارزمية على اكتشاف أوجه التشابه بين المدخلات، بحيث يتم تجميع العناصر التي تشترك في خصائص معينة ضمن مجموعات واحدة. وبعبارة أخرى، لا يمتلك النظام مخرجات محددة مسبقًا، لكنه يقوم باستكشاف البيانات واستخلاص استنتاجات منها بهدف وصف البُنى الخفية داخل البيانات غير المعنونة.[1]

و يعتمد التعلم الغير خاضع للإشراف على التجميع أو التحليل العنقودي يهدف إلى جمع مجموعة من العناصر بحيث تكون العناصر داخل كل مجموعة تسمى هذه المجموعة بالعنقود أكثر تشابهًا فيما بينها مقارنة بالمجموعات الأخرى. وبصياغة أدق، فإن الهدف من تحليل العناقيد هو تكوين مجموعات تكون فيها العناصر داخل كل مجموعة متشابهة قدر الإمكان، بينما تختلف هذه المجموعات عن بعضها بشكل واضح. ومن أهم خوارزميات التجميع نجد خوارزمية K-means وخوارزمية التجميع الهرمي.[2]

الشكل: شكل التجميع



المصدر: من اعداد الطالب

[1] ميلاد وزان ،**كتاب تعلم الالة و علم البيانات** ،ص143
[2] ميلاد وزان ،المرجع السابق ،ص328

تستخدم خوارزميات التجميع في:[1]

– تقسيم السوق حيث تقوم خوارزميات التجميع بتقسيم العملاء لعدة فئات بناء على الخصائص المتشابهة لكل فئة.

– الكشف عن السلوك غير الطبيعي بين العينات

– تحليل وتسمية البيانات الجديدة حيث تعمل خوارزميات التجميع على كشف الانماط داخل البيانات و ينشا لنا ذلك اهداف و مسميات اصطناعية بعد ان كنا نجهل الهدف من البيانات.

و ايضا تستخدم في:[2]

– التحليل المصرفي حيث تعمل خوارزميات التجميع على اكتشاف المعاملات المالية المشبوهة ووضع علامات عليها و ترميزها حيث ساعد هذا الاستخدام على التصدي بشكل فعال لظاهرة غسيل الاموال.

– المجال الطبي، حيث أثبتت خوارزميات التجميع جدارتها في هذا المجال، فهي تبسط عملية التعرف على أمراض العيون و مختلف الامراض الاخرى، حيث اعتبر هذا نموذجا كفؤا للغاية و ساهمت في تحسين دقة التشخيص الطبي بشكل غير مسبوق.

3- الشبكة العصبية الاصطناعية:

تعد الشبكات العصبية الاصطناعية من أبرز الأساليب الحديثة والمتقدمة في مجال الذكاء الاصطناعي، إذ تقوم على محاكاة البنية العصبية الموجودة في الخلايا العصبية داخل الدماغ البشري. وتهدف هذه الشبكات إلى تقليد الذكاء والسلوك الإنساني، من خلال القدرة على اكتساب المعلومات وتنظيمها بطريقة مشابهة للإنسان. وهي تتكوّن في الأساس من مجموعة وحدات مترابطة تُعرف بالعصبونات.[3]

[1] محمد لحلح، كتاب مدخل إلى الذكاء الاصطناعي وتعلم الآلة،2020، ص77

[2] Pritika Talwar, Shubham, and Komalpreet Kaur, **EXPLORING CLUSTERING TECHNIQUES IN MACHINE LEARNING**, International Journal of Creative Research Thoughts, 2024, p768

[3] كردودي سهام،بن قدور علي،مقال **التنبؤ بالشبكات العصبية االصطناعية كدعامة للمراجعة التحليلية في عملية التدقيق**،مجلة مجلة الإستراتيجية والتنمية، حالة مركب تكرير الملح SEL.NA.E (بسكرة) في الفترة 2010-2014،ص170

امثلة حول استخداماتها:[1]

تضمن معالجة اللغات الطبيعية (NLP) قراءة وفهم اللغة المنطوقة أو المكتوبة من خلال وسيط حاسوبي. وبالتالي، تساعد تقنيات معالجة اللغات الطبيعية الحاسوب على قراءة النصوص، وسماع الكلام، وتفسيره، وتحليل المشاعر، وتحديد الجوانب الهامة، حيث يمكن استخدام تقنيات الشبكة العصبية في هذا الصدد من الأمثلة على المهام المرتبطة بمعالجة اللغات الطبيعية تتمثل في المساعد الشخصي الافتراضي و روبوتات الدردشة التعرف على الكلام، وصف الوثائق، والترجمة اللغوية أو الآلية.

التعرف على الصور هو مثال معروف وواسع الانتشار للتعلم الآلي و الشبكات العصبية تحديدا في العالم الحقيقي، والذي يمكنه تحديد كائن ما كصورة رقمية. ومن الأمثلة الشائعة على ذلك :تصنيف الأشعة السينية كسرطانية أو غير ذلك، التعرف على الحروف و كشف الوجوه في الصور.

المطلب الثالث: العلاقة بين تعلم الآلة و ادارة الموارد البشرية:

هناك علاقة كبيرة بين تعلم الآلة و ادارة الموارد البشرية رغم ان تقنية تعلم الآلة مازالت حديثة في مجال ادارة الموارد البشرية الا انها تعتبر اداة مهمة جدا و تتكامل بشكل كبير في ادارة الموارد الشرية و سنتطرق الى استخدامات تعلم الآلة في ادارة الموارد البشرية و ايضا الى اهم المخاطر والاعتبارات الأخلاقية والامتثال لهذه التقنية داخل ادارة الموارد البشرية.

اولا: استخدامات تعلم الآلة في ادارة الموارد البشرية:

تتمثل الاستخدامات في:[2]

أ- التنبؤ بمغادرة الموظفين: يشكّل انفكاك الموظفين عن العمل أحد التحديات الجوهرية التي تواجه المنظمات نظير ما يترتب عليه من أعباء مالية باهظة، حيث تبرز الأهمية الاستراتيجية لامتلاك القدرة على التنبؤ باحتمالية المغادرة لصياغة سياسات استباقية للاستبقاء .ويُمكن تفعيل هذا التنبؤ منهجيًا عبر خوارزميات التعلم الآلي كالتصنيف والتجميع، وتحديدًا أسلوب التصنيف الثنائي

[1] Iqbal H. Sarker, **Machine Learning: Algorithms, Real-World Applications and Research Directions**, SN Computer Science, 2021, p160

[2] Rodrigo Miranda Cabrera , Manuel Leon Ganchozo , Rodríguez Salazar Oswaldo Rene , Rafael Wilfredo Rojas Bujaico , Héctor Huamán Samaniego , John Fredy Rojas Bujaico, **Impact Of Machine Learning In Human Resource Management: Towards The Modernization Of Leadership**, Journal of Positive School Psychology, 2022, p291, 292, 293

لفرز الموظفين إلى فئتين: المغادرين والمستمرين، بغية استباق الظاهرة وتلافي التكاليف المرتفعة المرتبطة بإحلال الكفاءات البديلة.

ب- تجزئة الموظفين : مثل تصنيف العملاء الى عدة فئات يمكننا ايضا تصنيف الموارد البشرية وفق محددات ديموغرافية، ومستويات أداء، وغيرها من المعايير التنظيمية. وتُطبّق هذه التجزئة عبر نماذج التعلم الآلي مثل خوارزمية K-means لإنشاء مجموعات عنقودية بالاستناد إلى السمات المخزنة في قواعد البيانات .وتستثمر المنظمات هذا التقسيم لتشخيص المتطلبات النوعية لكل فئة، مما يتيح تلبية احتياجاتهم الفكرية والعملية الفعلية، ويدعم كفاءة استراتيجيات الاستبقاء الوظيفي.

ج- الاستقطاب والتوظيف : يمثّل وضع الفرد المناسب في الموقع الوظيفي الملائم ركيزة حاسمة في إدارة الموارد البشرية، وتتجه مؤسسات رائدة مثل (Citigroup) و (Morgan Stanley)و (Goldman Sachs) نحو توظيف التعلم الآلي للتنبؤ بالأداء المستقبلي للمرشحين، حيث يضمن الاعتماد على هذه الأنظمة الذكية رفع دقة وملاءمة قرارات الانتقاء .حيث تستخدم نماذج التعلم الآلي في فحص السير الذاتية، إذ تتيح نماذج استخراج قواعد الارتباط (Association Rule Mining) ربط السير الذاتية للمترشحين بالمتطلبات المهنية المطلوبة، مما يعزز جودة مخرجات التوظيف . وتعتمد آلية استخراج قواعد الارتباط على التنبؤ بظهور متغير أو عنصر معين بناءا على توفر محددات أخرى، وفي سياق الفرز، تمكّن هذه التقنية من استشراف أداء المتقدمين استنادًا إلى امتلاكهم لمهارات تخصصية دقيقة. علاوة على ذلك، تُسهم هذه الخوارزمية في تمكين منصات التوظيف الرقمية من تقديم مقترحات مهنية ملائمة للباحثين عن عمل، عبر تحليل الارتباطات بين الفرص المتاحة والمسلك البحثي للمستخدم، فضلاً عن تتبع التفاعلات والمنشورات الصادرة عنه.

ثانيا: تحديات استخدام تعلم الآلة في ادارة الموارد البشرية

تعمل تقنية تعلّم الآلة في الموارد البشرية على توفير العديد من المزايا القوية، لكن في المقابل هناك تكاليف لهذه التقنية. ولتحقيق اكبر قدر من الاستفادة من هذه التقنية، يجب على إدارة الموارد البشرية تجنّب بعض المخاطر المحتملة وضمان شفافية و عدالة و توافق الانظمة مع القوانين.

ومن أهم الاعتبارات ما يلي:[1]

أ- جودة البيانات وإتاحتها :تعتمد خوارزميات التعلم الآلي على مدخلات بيانية دقيقة ومتنوعة لضمان كفاءة الأداء، غير أن بيانات الموارد البشرية الموزعة عبر الأنظمة قد تتضمن معلومات غير دقيقة أو متقادمة، مما يجعل التحقق من سلامة البيانات وتوافرها المستمر أحد أبرز التحديات المنهجية.

ب- التحيز والخصوصية المعلوماتية :إن النماذج الحوسبية القائمة على التعلم الآلي والتي يجري تدريبها ببيانات تاريخية للموارد البشرية قد تنتج مخرجات متحيزة، وهو أمر بالغ الأهمية في سياقات التوظيف، الاستقطاب، وتقييم الأداء. علاوة على ذلك، فإن الطبيعة السرية لبيانات العاملين الحاوية على مؤشرات حساسة تحتم تفعيل آليات الحماية والامتثال الصارم للأطر التشريعية المنظمة.

ج- إدارة التغيير التنظيمي :إذ يرفض بعض الموظفين و متخصصي ادارة الموارد البشرية المتشككين في القرارات الخوارزمية اعتماد الآليات التشغيلية القائمة على التعلم الآلي. ومن ثمّ، تستلزم المرحلة استثمارًا في استراتيجيات إدارة التغيير لضمان تحول تنظيمي مرن وتحقيق قبول أوسع.

د- شُح المهارات التخصصية :يتطلب دمج التعلم الآلي في إدارة الموارد البشرية كفاءات تمتلك مهارات تطوير النماذج وصيانتها وتفسيرها، بيد أن التنافسية ونقص الموارد يعيقان تأهيل الكوادر الداخلية أو استقطاب خبراء البيانات المتخصصين.

هـ- التكامل التقني والبنيوي :يسفر دمج تقنيات التعلم الآلي ضمن الأنظمة الرقمية المتقادمة عن ضرورة إجراء تعديلات جوهرية في البنية التحتية والمسارات الإجرائية، الأمر الذي قد يولد إرجاءً تنفيذيًا وممانعة من بعض الأطراف الفاعلة.

المبحث الثالث: الدراسات السابقة

بعد ان تطرقنا لاهم المفاهيم حول موضوع سلوك الموظفين و ايضا مفاهيم تعلم الآلة و ابرز استخداماته سنتطرق في هذا المبحث حول ابرز الدراسات السابقة و التي تطرقت لموضوع سلوك الموظفين او بالأخص خروج الموظفين و تسربهم وكذلك موضوع اساليب التحليل المختلفة و التي تتراوح بين الاساليب الوصفية و الاساليب الكمية.

[1] Deepti Sharma, Sellvasundaram, Prasanta Chatterjee Biswas, Suresh Kumar, Ganesh, and Kinjal Parmar, **Machine Learning and HRM: A Path to Efficient Workforce Management**, Journal of Informatics Education and Research, 2024, p796, 797

المطلب الأول: الدراسات العربية

الدراسة الأولى: دراسة مصعب حمد هويمل المسيعدين و أحمد نهار الرفوع (2021) تحت عنوان "العوامل المؤثرة في نية ترك العمل لدى العاملين في قطاع الصناعات البلاستيكية الأردنية"، هدفت هذه الدراسة إلى معرفة مدى تأثير العوامل عدالة التعويضات، ظروف وبيئة العمل، توفر وسائل السلامة العامة، وعدالة تقييم الأداء في نية ترك العمل من وجهة نظر العاملين في قطاع الصناعات البلاستيكية في الأردن. وتم أخذ عينة عشوائية من مجتمع الدراسة تكونت من (240) موظفاً من أصل (800) موظفاً. وكانت الاستبانة أداة للدراسة بعد التأكد من صدقها وثباتها وتكونت من (38) فقرة تقيس متغيرات الدراسة. وتم تحليلها باستخدام برنامج (SPSS 25) وباستخدام المنهج الوصفي التحليلي، بالإضافة إلى تحليل التباين وتحليل الانحدار لاختبار فرضيات الدراسة. وقد اشارت النتائج إلى وجود أثر ذو دلالة إحصائية للعوامل: (عدالة) التعويضات، ظروف وبيئة العمل، توفر وسائل السلامة العامة، عدالة تقييم الأداء في التقليل من نية ترك العمل حيث كان بعد توفر وسائل السلامة العامة الأكثر تأثيراً في نية ترك العمل، يليه بعد عدالة تقييم الأداء. وأوصت الدراسة بتوفير وسائل السلامة العامة لضمان بيئة مريحة وآمنة، كما أوصت بتهيئة ظروف وبيئة العمل المناسبة، وكذلك ضمان العدالة في تقييم الأداء.

الدراسة الثانية: دراسة حمزة معمري (2016) تحت عنوان "دور محددات الرضا الوظيفي في التنبؤ بمستوى الالتزام التنظيمي في المؤسسات"، هدفت هذه الدراسة إلى تحديد دور عدة عوامل مؤثرة في الرضا الوظيفي، وهي (الراتب، والترقية، والإشراف، والعلاقات الاجتماعية، والظروف المادية)، في التنبؤ بدرجة الالتزام التنظيمي لدى عينة مكونة من 156 موظفًا في شركة سونلغاز بمدينة ورقلة. وقد أظهرت الدراسة وجود علاقة إيجابية بين الالتزام والعوامل المذكورة، وأن تبني الشركات الجزائرية موقفًا إيجابيًا تجاه هذه العوامل يقلل من حالات اللامبالاة الإدارية ودوران الموظفين.

الدراسة الثالثة: دراسة النخلو عبد الوياب و عزري سيف الدين تحت عنوان "أثر العوامل التنظيمية على نية ترك العمل دراسة حالة لعينة من موظفي مؤسسة النقل الحضري وشبه حضري بورقلة"، هذه الدراسة عبارة عن مذكرة مقدمة لاستكمال متطلبات شهادة الدراسات الاكاديمية من جامعة قاصدي مرباح ورقلة 2022. تهدف هذه الدراسة إلى التعرف بشكل دقيق على العوامل التنظيمية و على نية ترك العمل، بالإضافة إلى دراسة العلاقة بين متغيرات الدراسة، وتم استخدام المنهج الوصفي التحليلي لدراسة الموضوع، كما تم الاعتماد على دراسة الحالة من خلال الاستبيان ومعالجته الإحصائية عن طريق برنامج spss نسخة 25، وتم توزيع الاستبيان على (65) موظف في مؤسسة النقل الحضري وشبه حضري بورقلة ، حيث توصلت الدراسة، بأن اهتمام المؤسسة بالناحية التنظيمية اهتمام متوسط

ويعتبر غير كافي، لا توجد فروق بين إجابات المبحوثين فيما يتعلق الجنس العمر، الخبرة، العلاقة الارتباطية بين العوامل التنظيمية.

المطلب الثاني: الدراسات الاجنبية

الدراسة الاولى: دراسة Rajkumar Govindarajan و Komal Kumar و Sudhakar Reddy و Sai Pravallika و Dhatri و Pavan Kumar هذه الدراسة عبارة عن ورقة بحثية بعنوان " Predicting Employee Attrition: A Comparative Analysis of Machine Learning Models Using the Analytics Dataset IBM Human Resource " حيث نشرت في مجلة elsevier عام 2025.

تبحث هذه الدراسة في أداء عدد من خوارزميات تعلّم الآلة في مهام التصنيف على بيانات منشاة من قبل شركة IBM حيث المتغير المدروس فيها هو مغادرة الموظفين، مع التركيز على دقتها في التنبؤ بالنتائج اعتمادًا على مجموعة بيانات معيّنة. تم تقييم خمسة نماذج: الغابة العشوائية (Random Forest)، وآلة دعم المتجهات (SVM) ، وشجرة القرار(Decision Tree) ، والتعزيز التدرّجي (Gradient Boosting)، ونموذج هجين يدمج عدة خوارزميات. وكان الهدف تحديد النموذج الذي يحقق أعلى دقة، وفهم نقاط القوة والضعف لكل نهج . أظهرت النتائج أن النموذج الهجين حقق أعلى دقة بلغت 88.7%، مما يشير إلى أن دمج خوارزميات مختلفة يساهم في الاستفادة من نقاط قوتها وتقليل نقاط ضعفها الفردية. وحققت آلة دعم المتجهات دقة بلغت 88.6%، مما يبرز قدرتها على التعامل مع حدود القرار المعقدة في الفضاءات عالية الأبعاد. كما حقق كلٌّ من نموذج الغابة العشوائية ونموذج التعزيز التدرّجي دقة بلغت 87.3%، وهو ما يعكس فعالية تقنيات التجميع (Ensemble) في تحسين الأداء التنبؤي من خلال تقليل الإفراط في التعلم (Overfitting) وتحسين تقليل الخطأ. في المقابل، أظهر مصنّف شجرة القرار أقل دقة بنسبة 80.5%، مما يبرز قابليته للإفراط في التكيّف وضعف قدرته على التعميم .يشير الأداء المتفوق للنموذج الهجين إلى توجه واعد للبحوث المستقبلية.

الدراسة الثانية: دراسة Hemanth Kumar و Meena و Santhanalakshmi و Chikati Srinu و Pushp Lata Rajpoot و Sunny Prakash هذه الدراسة عبارة عن مقال مفتوح بعنوان " Machine Learning and HRM: A Path to Efficient Workforce Management" حيث نشرت في مجلة Journal of Informatics Education and Research عام 2024.

تناولت هذه الدراسة كيفية استخدام خوارزميات التعلم الآلي لتحقيق أفضل ممارسات إدارة الموارد البشرية، لا سيما تلك المتعلقة بمهام مثل تحديد أداء الموظفين، وتحليل معدل دوران الموظفين،

وتخطيط القوى العاملة. وقد تم تقييم أداء خوارزميتي الغابة العشوائية والتعزيز التدرجي في أنشطة الموارد البشرية من خلال تحليل مجموعات بيانات واقعية خلال التجارب التي أُجريت. أظهرت البيانات المُبلغ عنها أعلى دقة لخوارزمية التعزيز التدرجي في التنبؤ بمعدل الشواغر باستخدام مقياسي الأداء MAE وRMSE، مما يشير إلى أفضل النتائج. وبالتالي، تتفوق خوارزمية التعزيز التدرجي على خوارزمية الغابة العشوائية في دقة التنبؤ، والضبط، والاستدعاء، ومقياس F1 في تحليل معدل دوران الموظفين حيث حققت الغابة العشوائية دقة 85% اما خوارزمية التدرج المعزز فقد حققت دقة بمقدار 87%. أما في تخطيط القوى العاملة، فقد أظهرت خوارزمية التعزيز التدرجي قيمًا أقل بقليل في متوسط الخطأ المطلق النسبي (MAPE) وجذر متوسط مربع الخطأ النسبي (RMSPE) مقارنةً بالنهجين الآخرين، مما يدل على تفوق أساليب التنبؤ. تُظهر النتائج المذكورة أعلاه بوضوح أن خوارزميات التعلم الآلي يمكن أن تكون ناجحة للغاية في مهام إدارة الموارد البشرية، وأن اتخاذ القرارات بناءً على البيانات يمكن أن يحسن الأداء ويعززه بشكل كبير.

الدراسة الثالثة: دراسة Ali Raza و Kashif Munir و Mubarak Almutairi و Faizan Younas و Mian Muhammad Sadiq Fareed هذه الدراسة عبارة عن مقال بعنوان" Predicting Employee Attrition Using Machine Learning Approaches" حيث نشرت على مجلة MDPI عام 2022

تهدف هذه الدراسة البحثية إلى تحليل العوامل التنظيمية التي تؤدي إلى تسرب الموظفين، بالإضافة إلى التنبؤ بتسرب الموظفين باستخدام تقنيات تعلم الآلة .وقد تم تطبيق أربع تقنيات من تقنيات تعلم الآلة بهدف المقارنة بينها. وحقق النموذج المقترح والمحسن Extra Trees Classifier دقة بلغت 93٪ في التنبؤ بتسرب الموظفين، متفوقًا بذلك على الدراسات الحديثة المتقدمة في هذا المجال. كما تم تطبيق تحليل البيانات الاستكشافي للموظفين (EEDA) لتحديد العوامل التي تؤدي إلى تسرب الموظفين. وكشفت الدراسة أن الدخل الشهري، والأجر بالساعة، والمستوى الوظيفي، والعمر تعد من أهم العوامل المؤثرة في تسرب الموظفين. وتساعد النتائج التي توصلت إليها هذه الدراسة والطريقة المقترحة المؤسسات على الحد من تسرب الموظفين من خلال تحسين العوامل التي تؤدي إليه.

الدراسة الرابعة: دراسة Dinh Cuong Nguyen و Gregory Lovisolo و Dan Tenney هذه الدراسة عبارة عم ورقة بحثية بعنوان " Employee Attrition: Leveraging Machine Learning Explainability to Identify Key Predictors of Turnover" نشرت في مؤتمر ASEE Northeast Section Conference عام 2025، تستخدم هذه الدراسة مجموعة بيانات موظفي اي بي ام، والتي تتضمن خصائص الموظفين مثل الرضا الوظيفي، والمشاركة في العمل، والتوازن بين العمل والحياة.

وقد تم استخدام عدة نماذج من تعلم الآلة للتنبؤ بتسرب الموظفين، من بينها الغابات العشوائية (Random Forests) وأشجار القرار (Decision Trees) والانحدار اللوجستي (Logistic Regression). وتم تقييم هذه النماذج باستخدام عدة مقاييس مثل الدقة (Accuracy) والدقة الإيجابية (Precision) والاستدعاء (Recall) ومنحنى AUC إضافة إلى مصفوفات الالتباس (Confusion Matrices). ومع ذلك، وعلى عكس الأساليب التقليدية، يدمج هذا البحث تقنيات MLX (التعلم الآلي القابل للتفسير) لتقديم فهم أعمق لسلوك النماذج.

اظهرت نتائج هذا الدراسة أن الرضا الوظيفي هو أقوى عامل للتنبؤ بتسرب الموظفين، يليه المشاركة في العمل ثم التوازن بين العمل والحياة. وقد حققت النماذج التنبؤية دقة بلغت 87.7٪، حيث قدم نموذج الغابات العشوائية (Random Forest) أكثر التنبؤات موثوقية. ومع ذلك، أظهرت إضافة تقنيات MLX (التعلم الآلي القابل للتفسير) أنه على الرغم من دقة النماذج، فإنها تفتقر إلى الشفافية الكاملة عند التنبؤ بتسرب موظفين محددين، مما يبرز الحاجة إلى التحسين المستمر للنماذج وتوفير تفسيرات أوضح لنتائجها.

الدراسة الخامسة: دراسة Ana Živković و Dario Šebalj و Jelena Franjković هذه الدراسة عبارة عن ورقة بحثية بعنوان "Prediction of the Employee Turnover Intention Using Decision Trees" نشرت في Proceedings of the 26th International Conference on Enterprise Information Systems عام 2024

تفحص هذه الدراسة فعالية منهجية شجرة القرار (Decision Tree) في التنبؤ بنية دوران الموظفين (Employee Turnover Intention)، وهو مجال حصلت فيه هذه الطريقة على قدر محدود من البحث. في هذه الورقة، تم إجراء بحث أولي وتم تطبيق أربع خوارزميات لشجرة القرار على عينة مكونة من 511 مستجيبًا. تدمج الدراسة عدة متغيرات تنبؤية في النموذج، بما في ذلك الرضا الوظيفي، والالتزام التنظيمي المدرك، والعدالة التنظيمية المدركة، والدعم التنظيمي المدرك، وفرص العمل البديلة المدركة، وذلك لتقييم تأثيرها على نية الدوران الوظيفي. وكان مقياس تقييم النموذج هو الاستدعاء(Recall) تشير النتائج إلى أن نموذج شجرة القرار باستخدام خوارزمية RandomTree كان ناجحًا نسبيًا في التنبؤ بنيات الدوران الوظيفي (بمعدل دقة يقارب 60٪)، حيث كان الرضا الوظيفي، وخاصة فرص النمو الشخصي والالتزام التنظيمي العاطفي، من المتنبئات المهمة. وتشمل العوامل المؤثرة الأخرى الرضا عن الراتب وعن الوظيفة نفسها، بالإضافة إلى العلاقات بين الأفراد. وتؤكد هذه الدراسة إمكانات طريقة شجرة القرار في إدارة الموارد البشرية وتوفر أساسًا للبحوث المستقبلية حول دور التحليلات التنبؤية في فهم ديناميكيات دوران الموظفين.

المطلب الثالث: التعقيب على الدراسات السابقة

من خلال مختلف الدراسات السابقة التي تطرقنا إليها, نلاحظ أن معظمها قد تطرق لموضوع مغادرة و خروج الموظفين, تنوعت هذه الدراسات السابقة بين الدراسات العربية و الدراسات الاجنبية, وقد اتفقت معظم هذه الدراسات على وجود محددات و عوامل تؤدي لترك او مغادرة الموظف المؤسسة حيث هناك من يعتبر ان الرضا الوظيفي من اهم العوامل و هناك من يرى ان العوامل المادية هي اهم العوامل ايضا، وكما نلاحظ ان المنهج الوصفي التحليلي كان هو السائد في الدراسات السابقة و هذا يدل على تحول مجال ادارة الاعمال و ادارة الموارد البشرية نحو الجانب التطبيقي اكثر، كما أن الدراسات التي عرضناها تعتبر حديثة ما بين (2019– 2025) و هذا ما يثبت ان موضوع دراستنا يعتبر حديثا و مواكبا لمختلف التغيرات التي تطرا على عالم ادارة الاعمال و مجال ادارة الموارد البشرية و ستتكلم على اوجه التشابه و اوجه الاختلاف لنرى الفرق بين الدراسة الحالية و الدراسات السابقة فيما يلي:

أوجه التشابه: اتفقت معظم هذه الدراسات على وجود محددات و عوامل تؤدي لترك او مغادرة الموظف المؤسسة، و أيضا استخدام تقنية التنبؤ خاصة الدراسات الاجنبية تحديدا.

أوجه الاختلاف: تقريبا كل الدراسات العربية اما كانت وصفية او استخدمت برامج تحليل جاهزة و قديمة و لم توظف تقنيات تعلم الآلة من خلال مكتبة scikit-learn من خلال البرمجة.

و من خلال هذا نلاحظ ان ما يميز دراستنا كونها قد تكون اول دراسة عربية تتطرق حول موضوع استخدام تقنية تعلم الآلة من خلال البرمجة و الادوات الحديثة.

خلاصة الفصل الأول:

في هذا الفصل تطرقنا للمفاهيم الاساسية و النظرية حول سلوك الموظفين حيث بينا اهميته و محدداته و ايضا تطرقنا لمفاهيم تعلم الآلة و بينا مدى ضرورته في كل المجالات و بالأخص مجال الموارد البشرية و التنبؤ بسلوك الموظفين، كما تم ذكر بعض الدراسات السابقة العربية المتنوعة حيث هناك من تكلمت على الموارد البشرية بصفة عامة و هناك من تكلم حول نفس الموضوع الخاص بنا الا وهو التنبؤ بخروج الموظفين باستخدام تقنيات تعلم الآلة و يجدر بالذكر ان هناك دراسة استخدمت نفس قاعدة البيانات و هي تخص Rajkumar Govindarajana و Komal Kumara و Sudhakar Reddy و Sai Pravallika و Dhatri و Pavan Kumar التي كانت بعنوان: Predicting Employee Attrition: A Comparative Analysis of Machine Learning Models Using the IBM Human Resource Analytics Dataset حيث هدفنا هو اجراء مقارنة نتائج التي تحصلنا عليها في دراستنا الحالية بتلك الدراسة السابقة و المتحصل عليها من خلال نماذج تعلم الالة في الفصل التطبيقي.

الفصل التطبيقي

الفصل الثاني : الفصل التطبيقي

بناءًا على ما تم تناوله في الفصل الأول حول الأطر النظرية لسلوك الموظف ومحددات قراراته تجاه المؤسسة، ينتقل الفصل الثاني إلى الجانب التطبيقي، تهدف هذه المرحلة إلى تطوير نموذج تنبؤي متخصص في توقع قرارات الموظفين بالبقاء أو الاستقالة، وذلك عبر تحليل المتغيرات والخصائص الأكثر تأثيراً في بناء النموذج، سيعتمد هذا العمل على خوارزميات التعلم الآلي (الخاضعة للإشراف والتصنيف)، مستفيداً من تنوع البيانات المتاحة التي تمنحنا فهماً دقيقاً للسلوك الوظيفي. كما سيستعرض الفصل المسار المنهجي المتبع، بدءاً من تعريف البيانات ومصادرها والأدوات المستخدمة، وصولاً إلى مراحل بناء النموذج وتقييمه، وانتهاءً بإجراء مقارنة مع دراسة سابقة استخدمت البيانات ذاتها لتقييم كفاءة النتائج.

المبحث الأول: استكشاف بيانات الدراسة

سنتطرق في هذا المبحث حول اهم متغيرات قاعدة البيانات التي عملنا عليها و اهم الادوات التي تم استخدامها في تطبيق تعلم الآلة و ايضا حول خطوات بناء نموذج تعلم الآلة و الذي مر بخطوات متنوعة و متسلسلة.

المطلب الأول: التعريف بمتغيرات الدراسة ومصدر جمع البيانات

أولا: مصدر البيانات

تم الاعتماد في هذه الدراسة على قاعدة بيانات خاصة بشركة من شركات التكنولوجيا و البرمجيات اي بي ام تم تحميلها من موقع Kaggle وهو موقع خاص بنشر البيانات التابعة للشركات والمؤسسات أو الجامعات التي تقوم بنشر بياناتها بهدف البحث العلمي ومساعدة الباحثين في وضع مقترحات أو حلول لمشكلة معينة وايضا التعلم و التدرب في مجال تحليل البيانات .

رابط قاعدة البيانات :[1]

[1] https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

ثانيا: وصف البيانات

تتكون البيانات من 35 متغير تشمل هذه المتغيرات 19 متغير نوعي 16 متغير كمي، ويمثل آخر متغير Attrition المتغير التابع أو ما يسمى ب "Target Variable" ، أما باقي المتغيرات فهي متغيرات مستقلة و الجدول التالي يشرح المتغيرات التي تم الاعتماد عليها في الدراسة:

المتغير	نوع المتغير	التفسير
Age	كمي	عمر الموظف.
Attrition	نوعي	هل ترك الموظف العمل؟
BusinessTravel	نوعي	مدى تكرار سفر الموظف للعمل
DailyRate	كمي	الأجر اليومي للموظف.
Department	نوعي	القسم الذي يعمل فيه الموظف
DistanceFromHome	كمي	المسافة بين منزل الموظف ومقر العمل بالكيلومترات.
Education	نوعي	المستوى التعليمي .
EducationField	نوعي	تخصص الموظف الدراسي (موارد بشرية، طب، تسويق، إلخ).
EmployeeCount	كمي	عدد الموظفين
EmployeeNumber	نوعي	الرقم الوظيفي الفريد لكل موظف.
EnvironmentSatisfaction	نوعي	مستوى الرضا عن بيئة العمل
Gender	نوعي	جنس الموظف.
HourlyRate	كمي	أجر الموظف بالساعة.
JobInvolvement	نوعي	درجة الاندماج والمشاركة في العمل.
JobLevel	نوعي	المستوى الوظيفي .
JobRole	نوعي	المسمى الوظيفي
JobSatisfaction	نوعي	مستوى الرضا الوظيفي.
MaritalStatus	نوعي	الحالة الاجتماعية .
MonthlyIncome	كمي	الدخل الشهري (الراتب).
MonthlyRate	كمي	المعدل الشهري (قيمة تقييمية أخرى للأجر الشهري).
NumCompaniesWorked	كمي	عدد الشركات التي عمل بها الموظف قبل الشركة الحالية.

هل الموظف فوق سن 18؟	نوعي	**Over18**
هل يعمل الموظف ساعات إضافية؟	نوعي	**OverTime**
نسبة الزيادة في الراتب مقارنة بالعام الماضي.	كمي	**PercentSalaryHike**
تقييم الأداء السنوي للموظف.	نوعي	**PerformanceRating**
مستوى الرضا عن العلاقات الاجتماعية داخل العمل.	نوعي	**RelationshipSatisfaction**
ساعات العمل القياسية .	كمي	**StandardHours**
مستوى خيارات الأسهم الممنوحة للموظف في الشركة.	نوعي	**StockOptionLevel**
إجمالي عدد سنوات خبرة الموظف المهنية.	كمي	**TotalWorkingYears**
عدد الدورات التدريبية التي حضرها الموظف في العام الماضي.	كمي	**TrainingTimesLastYear**
التوازن بين الحياة المهنية والشخصية .	نوعي	**WorkLifeBalance**
عدد السنوات التي قضاها الموظف في هذه الشركة الحالية.	كمي	**YearsAtCompany**
عدد السنوات التي قضاها الموظف في منصبه الحالي.	كمي	**YearsInCurrentRole**
عدد السنوات منذ آخر ترقية حصل عليها الموظف.	كمي	**YearsSinceLastPromotion**
عدد السنوات التي عمل فيها الموظف تحت إدارة مديره الحالي.	كمي	**YearsWithCurrManager**

المصدر :من اعداد الطالب

المتغير التابع: Attrition يمثل حالة العميل تجاه الشركة هل سيغادر أم لا

المتغيرات المستقلة: هي كل المتغيرات المتبقية

المطلب الثاني: ادوات وبيئة التطوير المستخدمة في الدراسة

في هذه الدراسة سنتطرق لمجموعة من الادوات تتمثل في:

أولا: برمجة Python

هي لغة برمجة متعددة الاستخدامات تم انشائها عام 1991م حيث تعتبر حاليا اقوى لغة برمجة في العالم، تعتبر بايثون Python لغة برمجة مستخدمة بشكل كبير وواسع في الكثير من المجالات المختلفة مثل تطوير الويب و تحليل البيانات و الامن السيبيراني و الذكاء الاصطناعي و بناء نموذج تعلم الالة و ذلك بسبب احتوائها على مكتبات و ادوات متخصصة في تلك المجالات و ايضا نظرا لسهولتها الكبيرة تحظى هذه اللغة بدعم هائل من قبل المتخصصين في تلك المجالات الكثيرة بالاضافة ايضا انها تعمل على انظمة تشغيل متنوعة مثل Windows و Mac و Linux و ايضا بسبب مفتوحة المصدر يمكن تحميلها مجانا و التطبيق بها مباشرة دون دفع اي مبلغ. ; سنتطرق في هذه الدراسة على المكتبات التي يمكننا من خلالها انشاء نموذج تنبؤي و التي تتمثل فيما يلي:

ا-مكتبة Numpy:

تعتبر مكتبة NumPy الخاصة ببرمجية البايثون من اهم المكتبات التي نقوم من خلالها بالعمليات الحسابية في القيم الجدولية او المصفوفات حيث صممت هذه المكتبة خصيصا مع المصفوفات او المتجهات (مصفوفة احادية الابعاد) حيث يمكننا ضرب المصفوفات في بعضها البعض او حساب المتوسط الحسابي للمصفوفة او الانحراف المعياري او معرفة اصغر او اكبر قيمة للمصفوفة و تتميز هذه المكتبة بسرعتها الحسابية الفائقة مقارنة بالحساب اليدوي بلغة بايثون و تشكل الأساس الذي تعتمد عليه معظم مكتبات التعلم الآلي والتعلم العميق حيث تكون العمليات الحسابية في النماذج تعمل من خلال مكتبة Numpy خلف الكواليس.

ب-مكتبة Pandas:

تعتبر مكتبة Pandas الخاصة ببرمجية البايثون واحدة من اكثر الادوات و المكتبات شهرة في مجال تحليل البيانات و التنقيب عليها حيث تتعامل هذه المكتبة مع الجداول خصيصا سواء كان جدول احادي الابعاد (عمود) او جدول ثنائي الابعاد (عدة اعمدة معا) و ايضا يمكننا من خلال هذه المكتبة استرداد البيانات و تحليلها و استكشافها و معالجتها و تنظيفها من القيم المفقودة و القيم الشاذة و القيم المتكررة و تجهيزها لعملية بناء النموذج التنبؤي و يمكن لهذه المكتبة التعامل مع ملفات مختلفة مثل ملف Excel و CSV و Json .

ج-مكتبة Scikit-Learn:

تعتبر مكتبة Scikit-Learn من أهم مكتبات لغة Python في مجال تعلم الآلة، و الاكثر استخداما، حيث يمكن من خلال هذه المكتبة عمل مهام التنبؤ بتعلم الالة المتعارف عليها مثل التنبؤ بقيمة عددية مستمرة بواسطة الانحدار او التنبؤ بقيمة فئوية بواسطة التصنيف او اكتشاف الانماط المخفية و تجميع العينات الى الفئات التي تنتمي اليها بناءا على التشابه بواسطة التجميع و تسمح لنا هذه المكتبة ببناء النموذج التنبؤي بواسطة تقسيم البيانات لبيانات تدريب و بيانات اختبار و استدعاء الخوارزمية او النموذج الذي نريده ثم ندرب النموذج على بياناتت التدريب ثم نقوم بجعل النموذج بتنبؤ بقيمة جديدة بعدها تسمح لنا هذه المكتبة من تقييم اداء النموذج من خلال استدعاء مؤشرات التقييم الخاصة بنماذج تعلم الالة.

د-مكتبة Seaborn:

تعتبر مكتبة Seaborn من اهم مكتبات انشاء الرسوم البيانية و تصور البيانات حيث يمكننا من خلالها انشاء الرسومات البيانية الاحصائية المعقدة بسهولة اكبر و اسرع و بشكل اكثر جاذبية مثل رسم الاعمدة النسبية و نقاط انتشار العينات و ايضا منحنى الانحدار.

ثانيا: بيئة التطوير Google Colab

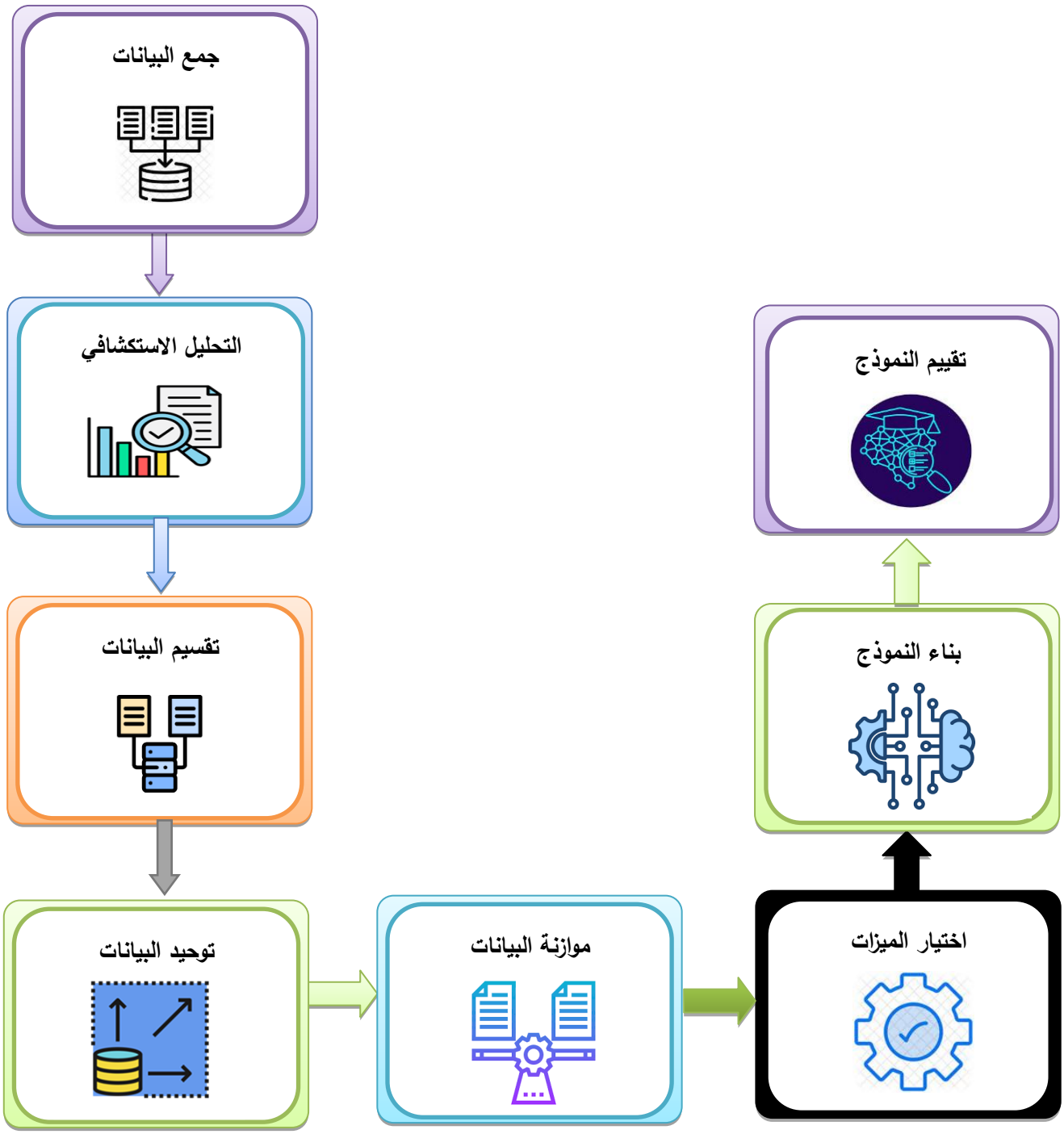
يعد Google Colab أداة قوية بشكل لا يصدق لتطوير مشاريع التعلم الآلي وتقديمها بشكل تفاعلي، حيث تعتبر المكان الذي يمكننا من خلاله كتابة اكواد لغة بايثون اضافة لتنفيذ مكتباتها المتطورة و القوية في هذه البيئة و ذلك بسبب دعمها الهائل لأدوات تحليل البيانات و يمكن من خلاله استرداد البيانات فيه و عمل التحليلات على تلك البيانات و مرورا بكل اجراءات مراحل بناء النموذج التنبؤي.

المطلب الثالث: خطوات بناء النموذج وأدوات تقييمه

بعد ان تعرفنا على البيانات الخاصة بالدراسة ومصدرها في المطلب الاول وايضا تعرفنا على ابرز الأدوات المستخدمة وبيئة التطوير التي سيتم الإعتماد عليها في إنشاء النموذج التنبؤي، حان الوقت الان لمعرفة ابرز المراحل الخاصة لهذا النموذج بسبب انها ضرورة منهجية من اجل شرح كيفية عمل النموذج بدءا من جمع البيانات و تنظيفها و معالجتها و موازنتها و اختبار المتغيرات المهمة وصولا الى تدريب النموذج و تقييمه دقته.

اولا :مراحل بناء النموذج

الشكل: خطوات بناء النموذج



المصدر: من اعداد الطالب

1- جمع البيانات:

هو الحصول على البيانات التي سوف نستخدمها في هذه الدراسة من اجل بناء نموذج للتنبؤ بمغادرة الموظفين، حيث كل الدراسات تقوم على البيانات التي يمكن جمعها بطرق متعددة و من مصادر كثيرة، و على سبيل المثال تم جمع بيانات هذه الدراسة من مصدر مفتوح و معروف و هو موقع Kaggle.

2- التحليل الاستكشافي للبيانات:

يعتبر من اهم المراحل حيث نتأكد من ان البيانات سليمة و خالية من الاخطاء و القيم المفقودة من اجل تنظيفها من خلال معالجة القيم المفقودة و القيم الشاذة و القيم المتكررة من خلال حذفها او تعويضها بقيمة اخرى من اجل ان تكون دقة النموذج على هذه البيانات دقيقة و سليمة.

3- تقسيم البيانات:

يعتبر مرحلة مهمة جدا حيث نقوم بتقسيم البيانات الى بيانات تدريب يتدرب عليها النموذج و بيانات اختبار نختبر بها النموذج.

4- توحيد البيانات:

هو عملية تحويل قيم المزايا (features) في مجموعة البيانات لتصبح ضمن نطاق محدد، مثل 0 إلى 1 أو −1 إلى 1. الهدف من هذه العملية هو ضمان عدم سيطرة ميزة واحدة على حسابات المسافة في الخوارزمية، مما يساعد في تحسين أداء الخوارزمية بشكل عام.

5- موازنة البيانات:

هي عملية ضبط توزيع الفئات في مجموعة البيانات لمعالجة مشكلات عدم توازن الفئات، حيث نستخدم الموازنة في حالة اذا كانت فئة اكثر بكثير من فئة اخرى داخل المتغير التابع.

6- اختيار الميزات:

يقصد بها اختيار المتغيرات المستقلة و التي تكون لها علاقة ارتباط قوية بالمتغير التابع و التي تعطينا اعلى دقة للنموذج مع تجنب المتغيرات المستقلة الغير مهمة و التي لها علاقة ضعيفة بالمتغير التابع و التي قد تفسد دقة النموذج.

7- بناء النموذج:

يقصد به اختيار الخوارزمية المناسبة واختيار افضل معلماتها و تدريبها على بيانات التدريب و اختيار النموذج الذي يعطي اعلى دقة ممكنة في عملية التنبؤ لهذه الدراسة.

8- تقييم النموذج:

يتم في هذه المرحلة التحقق من مدى دقة النموذج في التنبؤ والتأكد ان النموذج موثوق و يعتمد عليه في هذه الدراسة بالاعتماد على مؤشرات خاصة بتقييم النماذج التنبؤية مثل Confusion Matrix و Accuracyو Precision و Recall و F1 Score

ثانيا: أدوات تقييم النموذج

بعد تجهيز البيانات لبناء النموذج الخاص بالدراسة يجب أن نتعرف على بعض الأدوات التي سيتم من خلالها تحديد النموذج الأنسب والتي تعتبر كمؤشرات لاختيار أدق النماذج وهي:

1- مصفوفة الارتباك

تعتبر مصفوفة الالتباس اداة بسيطة و قوية في تقييم نموذج التعلم الالي الخاص بعمليات التصنيف حيث نعرف من خلال اين يكون النموذج قوي في التصنيف و اين يكون النموذج ضعيف في التصنيف و ايضا معرفة اين اخطا و اين اصاب و هذا يعتبر بالغ الاهمية في تقييم النموذج و يوضح الشكل الاتي المصفوفة:

جدول: مصفوفة الإرتباك Counfusion Matrix

	Actual Positive **الحقيقة الفعلية**	**Actual Negative** **السلبية الحقيقية**
Predicted Positive الايجابية المتنبؤة	True Positive(TP)	**False Positive(FP)**
Predicted Negative السلبية المتنبؤة	False Negative(FN)	**True Negative(TN)**

تتمثل الفئات الأربع لمصفوفة الارتباك في:

- **True Positive (TP):** وهي الإيجابية الصحيحة و هي عدد الحالات التي تنبأ بها النموذج النتيجة بانها ايجابية و النتيجة الفعلية كانت إيجابية بالفعل اي ان النموذج توقع النتيجة الايجابية بشكل صحيح تماما مثل (توقع تخلي الموظفين عن الوظيفة و قد تخلوا فعلا عن الوظيفة).

- **False Negative (FN):** وهي السلبية الخاطئة و هي عدد الحالات التي تنبأ بها النموذج النتيجة بانها سلبية لكن النتيجة الفعلية كانت إيجابية اي ان النموذج توقع النتيجة السلبية بشكل خاطئ مثل (توقع عدم تخلي الموظفين عن الوظيفة لكنهم تخلوا عن الوظيفة).

- **False Positive(FP) :** وهي الايجابية الخاطئة و هي عدد الحالات التي تنبأ بها النموذج النتيجة بانها ايجابية لكن النتيجة الفعلية كانت سلبية اي ان النموذج توقع النتيجة الايجابية بشكل خاطئ مثل (توقع تخلي الموظفين عن الوظيفة لكنهم لم يتخلوا عن الوظيفة).

- **True Negative(TN):** وهي السلبية الصحيحة و هي عدد الحالات التي تنبا بها النموذج النتيجة بانها سلبية و النتيجة الفعلية كانت سلبية فعلا اي ان النموذج توقع النتيجة السلبية بشكل صحيح تماما مثل (توقع عدم تخلي الموظفين عن الوظيفة ولم يتخلوا فعلا عن الوظيفة).

و تحتوي مصفوفة الارتباك على المؤشرات التالية:

ا- الدقة العامة (Accuracy) :

يوضح مؤشر Accuracy عدد التوقعات او التنبؤات الصحيحة التي تنبا بها النموذج في النتائج الايجابية و السلبية من بين إجمالي جميع التوقعات. حيث يعطي هذا المؤشر فكرة عن الأداء العام للنموذج. و يمكننا حساب هذا المؤشر بالمعادلة الاتية:

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

ب- الدقة التنبؤية (Precision) :

يوضح مؤشر Precision نسبة و جودة توقعات النموذج الإيجابية فهو يخبرنا عن عدد التوقعات التي توقع النموذج على أنها ايجابية وكانت صحيحة بالفعل في الواقع. و يمكن حسابه بالمعادلة الاتية:

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

ج- الاستدعاء (Recall) :

يوضح مؤشر Recall مدى قدرة النموذج على التقاط الحالات الإيجابية. و يمكن حساب المؤشر بالمعادلة الاتية:

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

د- مقياس (F1 Score) :

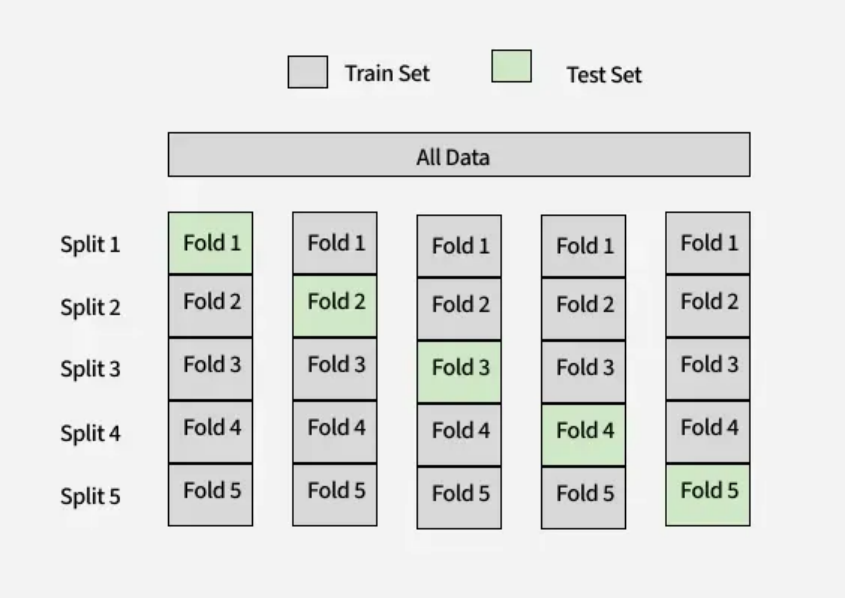
يقوم F1 Score بالجمع بين الدقة Precision والاستدعاء Recall في مؤشر واحد لتحقيق التوازن بينهما ،فهو يوفر رؤية أفضل و ادق لأداء النموذج العام، خاصة اذا كانت بياناتنا غير المتوازنة ،و يمكن حساب المؤشر بالمعادلة الاتية:

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

2- التحقق المتقاطع Cross Validation

يعتبر التحقق المتقاطع اداة قوية جدا في تقييم اداء نموذج تعلم الآلة حيث يستخدم في حالة عدم تأكدنا من ان النموذج سيعطي دقة عالية في حال مواجهة بيانات جديدة لم يراها من قبل. حيث يمنحنا نظرة مسبقة حول كيفية أداء النموذج في العالم و ايضا يساعدنا في التأكد من ان تنبؤاتنا ليست مجرد ضربة حظ بل تنبؤات موثوقة و يمكن تعميم النتائج على البيانات الجديدة و التي لم نرها من قبل. تتمثل آلية عمله في انه يقوم بتقسيم مجموعة البيانات على عدة أجزاء وتسمى بالطيات (Folds) . يتم تدريب النموذج على طيات التدريب ويختبر على طية الاختبار الأخيرة المتبقية. تتكرر هذه العملية عدة مرات بحسب عدد الطيات، بحيث تُستخدم كل طية كطقم اختبار لمرة واحدة بالضبط بعد ذلك يتم حساب متوسط أداء النموذج عبر جميع التكرارات لتقديم تقدير قوي ومتين لقدرته على التعميم، و يمكن شرح التحقق المتقاطع بشكل اوضح في الصورة الاتية:

الشكل: مصفوفة التحقق المتقاطع **Cross Validation**



المصدر: موقع https://www.geeksforgeeks.org

من خلال الصورة السابقة يمكننا توضيح مصفوفة التحقق المتقاطع بشكل أدق حيث نلاحظ أنه تم تقسيم البيانات على خمس طيات بحيث أنه في الدورة الاولى يتم إختبار الطية الأولى (Fold 1) وتدريب باقي الطيات (Fold2/ Fold3/ Fold4/ Fold5) ، أما في الدورة الثانية يتم إختبار الطية الثانية (Fold2) وتدريب باقي الطيات (Fold1/ Fold3/ Fold4/ Fold5) حيث اصبحت الطية الاولى في الدورة الثانية طية تدريب و الطية الثانية طية اختبار و نعيد هذه العملية الى ان نصل الى آخر دورة بحيث تصبح الطية الخامسة هي طية الاختبار و باقي الطيات طيات تدريب و بعدها نحسب متوسط اداء الدورات الخمس لنخرج بنتيجة التحقق المتقاطع.

المبحث الثاني : تحليل النتائج وتفسيرها

بعدما تطرقنا الى بيانات الدراسة، مصدر تجميعها، و اهم الأدوات التي سيتم استخدامها في بناء نموذجنا الخاص للتنبؤ بسلوك الموظفين و الذي سيتمثل في خروج الموظفين او بقائهم، وأيضا بعد ان تطرقنا الى أهم المراحل التي سيتم إعتمادها في بناء النموذج سوف نتطرق في هذا المبحث الى تحليل و تفسير نتائج قاعدة البيانات باستخدام لغة البرمجة Python و اهم مكتباتها و google colab

المطلب الأول: استكشاف البيانات و معالجتها

بعد الانتهاء من جمع البيانات، تأتي مرحلة تحليلها، غير أنه قبل الشروع في ذلك يجب التأكد من أن البيانات المعتمدة في بناء النموذج سليمة وخالية من القيم المفقودة أو البيانات المكررة التي قد تؤثر على دقته. ونظرًا لأن البيانات المستخدمة في هذه الدراسة تضم 35 متغيرًا و1470 عينة، فإنها قد تحتوي على قيم مفقودة أو بيانات غير متناسقة، مما يستوجب القيام بعملية تنظيف لها.

أولا: سنسترد المكتبات الخاصة ببايثون و التي يمكننا من خلالها عمل تحليل و معالجة قاعدة البيانات بشكل فعال و سلس و يمكننا ذلك من خلال الاوامر الاتي:

الشكل: إستيراد المكتبات

```
import pandas as pd
import numpy as np
import seaborn as sns
```

المصدر: لغة Python

ثانيا : إستيراد قاعدة البيانات التي سيتم العمل عليها من خلال الأمر الاتي:

الشكل: إستيراد قاعدة البيانات

```
df=pd.read_csv('/content/hr_analytic.csv',sep=';')
```

المصدر: لغة Python

ونظرا لأن قاعدة البيانات كبيرة لأنها تحتوي على 1470 عينة ولا يمكن حصرها في جدول واحد لذا سأظهر فقط أول 5 أسطر من البيانات باستخدام الأمر الآتي:

```
df.head()
```

	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	Employe
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	

5 rows × 35 columns

المصدر: لغة Python

ثالثا: بعد عرض البيانات المتعلقة بالدراسة من اجل عمل تحليل استكشافي عليها من اجل معرفة نوع المتغيرات وعددها وهل توجد بها قيم مفقودة أو لا بإستخدام الامر الاتي:

الشكل: امر استكشاف البيانات

```
df.info()
```

المصدر: من اعداد الطالب

فتظهر النتائج الموالية:

```
Data columns (total 35 columns):
#   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
0   Age                       1470 non-null   int64
1   Attrition                 1470 non-null   object
2   BusinessTravel            1470 non-null   object
3   DailyRate                 1470 non-null   int64
4   Department                1470 non-null   object
5   DistanceFromHome          1470 non-null   int64
6   Education                 1470 non-null   int64
7   EducationField            1470 non-null   object
8   EmployeeCount             1470 non-null   int64
9   EmployeeNumber            1470 non-null   int64
10  EnvironmentSatisfaction   1470 non-null   int64
11  Gender                    1470 non-null   object
```

```
12  HourlyRate                1470 non-null   int64
13  JobInvolvement            1470 non-null   int64
14  JobLevel                  1470 non-null   int64
15  JobRole                   1470 non-null   object
16  JobSatisfaction           1470 non-null   int64
17  MaritalStatus             1470 non-null   object
18  MonthlyIncome             1470 non-null   int64
19  MonthlyRate               1470 non-null   int64
20  NumCompaniesWorked        1470 non-null   int64
21  Over18                    1470 non-null   object
22  OverTime                  1470 non-null   object
23  PercentSalaryHike         1470 non-null   int64
24  PerformanceRating         1470 non-null   int64
25  RelationshipSatisfaction  1470 non-null   int64
26  StandardHours             1470 non-null   int64
27  StockOptionLevel          1470 non-null   int64
28  TotalWorkingYears         1470 non-null   int64
29  TrainingTimesLastYear     1470 non-null   int64
30  WorkLifeBalance           1470 non-null   int64
31  YearsAtCompany            1470 non-null   int64
32  YearsInCurrentRole        1470 non-null   int64
33  YearsSinceLastPromotion   1470 non-null   int64
34  YearsWithCurrManager      1470 non-null   int64
dtypes: int64(26), object(9)
 memory usage: 402.1+ KB
```

و منه نلاحظ ان بياناتنا نظيفة جدا و لا تحتوي على قيم مفقودة من خلال عرضها و سنقوم حاليا بعرض معلومات احصائية حول البيانات لرؤية نتائج استكشافها بشكل ادق و اسهل :

جدول: المعلومات الإحصائية لقاعدة البيانات قبل المعالجة

Number of variables	**35**
Number of observations	1470
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	402.1 MB

المصدر: من إعداد الطالب

نلاحظ من خلال الجدول السابق أنه لدينا : 35 متغير في قاعدة البيانات الكلية ،كما أن عدد العينات بلغ 1470 عينة ، وكان عدد القيم المفقودة 0 أي بنسبة مقدرة ب0.00% ، و ايضا نلاحظ عدم وجود أي قيم مكررة في قاعدة البيانات بنسبة مقدرة ب0.00% و منه قاعدة بياناتنا لا تحتاج لمعالجة و تنظيف اصلا بسبب سلامتها و عدم احتوائها على اي قيمة مفقودة او قيمة مكررة .

رابعا: تقسيم البيانات

بعد ان تأكدنا من بياناتنا نظيفة و سليمة يأتي الان دور تقسيم البيانات الى بيانات تدريب و التي سيتدرب عليها النموذج و بيانات اختبار و التي سنختبر بها دقة نموذجنا و ايضا درجة موثوقية نموذجنا من خلالها حيث قمنا بتقسيم البيانات الى 85 بالمئة بيانات تدريب و 15 بالمئة بيانات اختبار.

الشكل: امر تقسيم البيانات

```
x_train,x_test,y_train,y_test=train_test_split(x,y,random_state=42,test_size=0.15)
```

المصدر: من اعداد الطالب

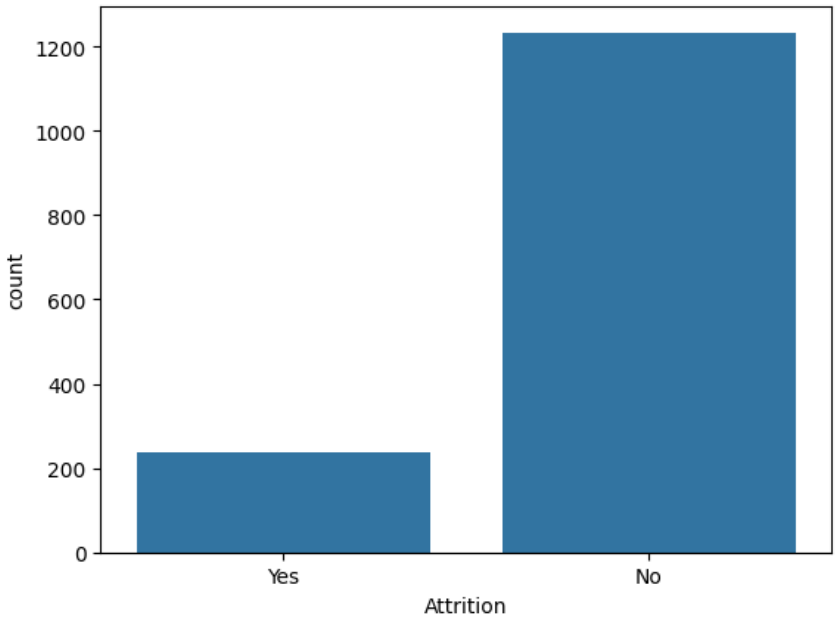
خامسا: تحجيم البيانات و ترميزها

حيث قمنا بتحويل قيم مزايا او اعمدة البيانات لتصبح ضمن نطاق 0 الى 1 لضمان عدم سيطرة ميزة على ميزة اخرى لكي لا تتشتت الخوارزميات المعتمدة و ايضا قمنا بترميز البيانات الفئوية و تحويلها لأرقام عددية من اجل ان يتمكن نموذجنا من حسابها و قراءتها.

سادسا: موازنة البيانات

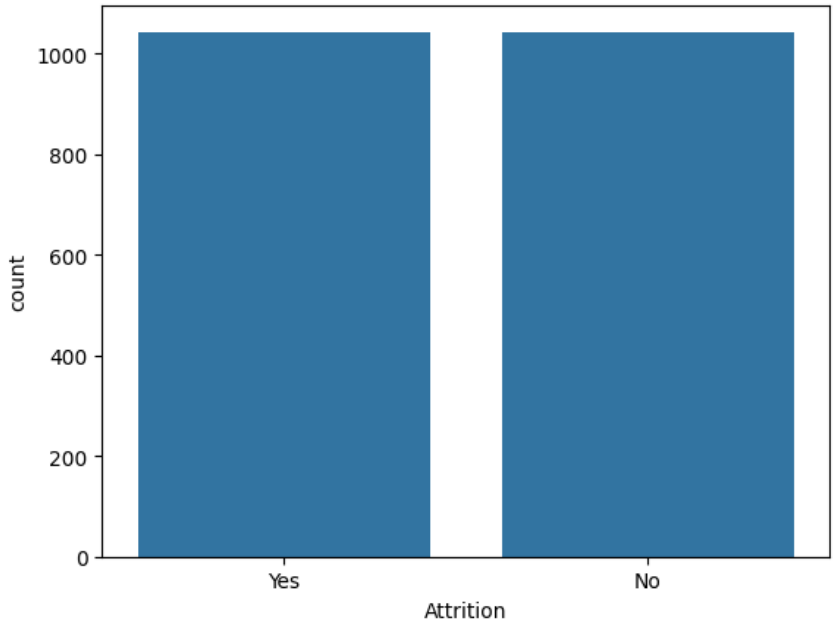
واجهنا في هذه البيانات مشكلة عدم توازن البيانات حيث ان فئة الموظفين الذين لم يغادروا كانت نسبتهم 83.51 بالمئة و فئة الموظفين الذين غادروا كانت نسبتهم 16.49 بالمئة حيث تعمل هذه المشكلة على جعل النموذج يركز على الفئة الاكثر و يهمش الفئة الاقل لذلك قمنا بموازنتها بواسطة تقنية التضخيم الاصطناعي لعينات الفئة الاقل لتتساوى مع عدد عينات الفئة الاكثر

الشكل: البيانات قبل الموازنة



المصدر: من اعداد الطالب

الشكل: البيانات بعد الموازنة



المصدر: من اعداد الطالب

سابعا: تحديد المتغيرات

بعد معالجة البيانات وترميزها و تحجيمها و موازنتها يجب ان نقوم بتحديد المتغيرات او الميزات الأكثر ارتباطا بالمتغير التابع وأقل ارتباطا في ما بينها من خلال تقنية تسمى التغليف Feature Sequential Selector وهي طريقة تندرج ضمن تقنيات التغليف wrapper حيث تحدد الميزات أو السمات المهمة والأكثر تأثيرا في بناء نموذج باستخدام تعلم الآلة حيث تعمل هذه التقنية على البدء بكل الميزات المتاحة و تقوم بالحذف التدريجي للميزات التي تسبب انخفاض دقة النموذج و ايضا تأخذ بعين الاعتبار تفاعل الميزات ببعضها بحيث يمكن ان تكونا ميزتان منفصلتان ضعيفتان في رفع دقة النموذج اما اذا عملا معا فيرفعان من الدقة و تتميز هذه الطريقة بدقتها العالية بالإضافة الى انها مرنة للغاية في التعامل مع كل نماذج و خوارزميات تعلم الآلة وتكمن أهمية إستعمال تقنية wrapper او بالاخص Selector Feature Sequential في كونها تساعد بشكل كبير على تحسين دقة التنبؤ وذلك بسبب أنها تختلف عن الطرق التقليدية التي تعتمد فقط على العلاقات الإحصائية بين المتغيرات و ايضا الطرق الاختيار اليدوية حيث لا يمكن اختيار الميزات يدويا اذا كان عددها كبيرا جدا كما أنها تساهم وتتخلص من الميزات غير المؤثرة في النموذج بسرعة عالية . و من خلال هذه التقنية تم اختيار الميزات التالية :

اولا : الميزات الفئوية الي تم ترميزها

حيث تخص الميزات التي تم ترميزها بواسطة OneHotEncoder() عليها :

ohe__BusinessTravel_Travel_Frequently

ohe__BusinessTravel_Travel_Rarely

ohe__EducationField_Life Sciences

ohe__EducationField_Marketing

ohe__EducationField_Medical

ohe__Gender_Male

ohe__JobRole_Laboratory Technician

ohe__JobRole_Research Scientist

ohe__JobRole_Sales Executive

ohe__JobRole_Sales Representative

ohe__MaritalStatus_Single

ohe__OverTime_Yes

ثانياً: الميزات الرقمية التي تم توحيدها

تخص البيانات الرقمية التي تم توحيدهما :

std__Age

std__DailyRate

std__DistanceFromHome

std__Education

std__EnvironmentSatisfaction

std__HourlyRate

std__JobSatisfaction

std__MonthlyIncome

std__NumCompaniesWorked

std__PercentSalaryHike

std__PerformanceRating

std__RelationshipSatisfaction

std__TrainingTimesLastYear

std__WorkLifeBalance

std__YearsInCurrentRole

std__YearsSinceLastPromotion

std__YearsWithCurrManager

المطلب الثاني : بناء و تقييم النموذج

بعد الانتهاء من تحديد المتغيرات المناسبة تأتي أهم مرحلة تهدف إليها هذه الدراسة وهي بناء النموذج الذي سيتنبأ بخروج الموظفين من المؤسسة او بقائهم في المؤسسة، بناء على الميزات التي تم تحديدها مسبقا سنقوم بتدريب عدة نماذج للتنبؤ والمقارنة بينها واختيار أفضل نموذج حيث تم استعمال هذه النماذج لتصنيف سلوك الموظفين نظرا لقدرتهم على التعامل مع البيانات الفئوية و العددية.

سوف يتم ذلك من خلال لغة Python و مكتبتها في تعلم الالة Scikit-learn

أولا : شجرة القرار Decision Tree

جرى الاعتماد على نموذج شجرة القرار، الذي يستند إلى آلية التقسيم المتتالي والتدريجي للبيانات بناءا على المتغيرات الأكثر تأثيرا، حيث يوظف مقياس الإنتروبي لتحديد المتغير الأمثل لقيادة عملية التقسيم في كل مرحلة. هذا ويمكن بناء وتدريب نموذج شجرة القرار برمجياً باستخدام الأمر التالي:

الشكل: استرداد نموذج شجرة القرار

```
from sklearn.tree import DecisionTreeClassifier
dt=DecisionTreeClassifier(random_state=42,criterion='entropy',max_depth=23)
```

المصدر: من اعداد الطالب

ا- تقييم أداء النموذج باستخدام مصفوفة الارتباك :

وظفنا مصفوفة الارتباك كأداة لتقييم النموذج في التنبؤ بمعدل دوران الموظفين، ونظراً لطبيعة الدراسة التي تندرج ضمن مسائل التصنيف الثنائي، كان لدينا فئتين فئة وتشمل الموظفين الذين غادروا المؤسسة فعلياً، والفئة الذين بقوا. وكان تنفيذ المصفوفة برمجيا بنموذج شجرة القرار كالاتي:

الشكل: الأمر الخاص باستخراج مصفوفة الارتباك

```
confusion_matrix(y_test,y_pred_dt)
```

المصدر: من اعداد الطالب

حيث كانت النتائج التي اظهرتها التي تخصنا عليها من خلال مصفوفة الالتباس لنموذج شجرة القرار كالتالي:

	Actual Positive الموظف خرج فعلا	Actual Negative الموظف بقي فعلا
Predicted Positive توقع خروج الموظف	9 (TP)	19(FP)
Predicted Negative توقع بقاء الموظف	22 (FN)	172 (TN)

المصدر: من إعداد الطالب

بناءا على مصفوفة الالتباس نلاحظ أن نموذج شجرة القرار توقع أن 9 (True Positive) موظفين سيخرجون من المؤسسة وقد خرجوا وأخطأ في توقع 19 موظف انهم سيخرجون من المؤسسة ولكن لم يخرجوا (False Positive) ، كما نلاحظ أن قيمة (False Negative) كانت 22 أي أنه هناك 22 موظف توقعهم النموذج بأنهم سيبقون لكنهم خرجوا و هذا يعني ان النموذج خطا اكثر مما اصاب في فئة الموظفين الذين سيخرجون و اصاب في172 موظف توقع النموذج بانهم سيبقون و بقوا فعلا (True Negative) .

ب- مؤشرات الأداء العامة

جدول: المؤشرات العامة حسب نموذج شجرة القرار

Accuracy	0.8190
Precision	0.3076
Recall	0.2580
F1 Score	0.2807

المصدر: من إعداد الطالب

بعد بناء نموذج شجرة القرار على بيانات الموظفين تم تقييم أدائه باستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 81,90% وهي نسبة قوية تدل على الأداء ممتاز بشكل عام أي أن النموذج تمكن من تصنيف ما نسبته 81,90% من البيانات بشكل صحيح، لكن قدرت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة 30,76% وهي نسبة ضعيفة في التصنيف الفئة الايجابية، و بلغت نسبة مؤشر الاسترجاع Recall ب25,80% في ادراك الموظفين الذين سيخرجون وهي نسبة ضعيفة وتعني أن النموذج استطاع الامساك ب25,80% من الموظفين الذين سيخرجون لكن افلت منه نسبة كبيرة من الموظفين الذين سيخرجون قدرت نسبتهم ب74,20% حيث توقعها على أنها قيم سلبية إلا أنها كانت قيم ايجابية و ذلك راجع لأنه توقع بشكل صحيح في الفئة الثانية و اخطأ كثيرا في الفئة الاولى ، و قدرت نسبة مؤشر F1 Score ب28,07% وهي نسبة ضعيفة تدل على أن النموذج لم يحقق توازنا بين مؤشر الدقة الايجابية Precision ومؤشر الإسترجاع Recall.

ج- تقييم أداء النموذج باستخدام مصفوفة التحقق المتقاطع Cross Validation

جدول: نتائج مصفوفة التحقق المتقاطع لنموذج شجرة القرار

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.9217	0.9220	0.8653	0.9860
Fold2	0.9327	0.9329	0.8739	1
Fold3	0.9287	0.9284	0.8669	1
Fold4	0.8977	0.8937	0.8143	1
Fold5	0.8899	0.8958	0.8738	0.9065
Mean cv	0.914	0.915	0.859	0.978

المصدر: من إعداد الطالب

نلاحظ من الجدول نتائج التحقق المتقاطع الخاص بنموذج شجرة القرار حيث نلاحظ نتائج النموذج اصبحت افضل بكثير مما هو عليه في التقسيمة الواحدة بعد ان عملنا علي التحقق المتقاطع و ذلك بسبب اختفاء التحيز في البيانات مما سمح للنموذج بالتعامل مع فئة الموظفين الذين خرجوا و الذين بقوا بشكل عادل و دقيق اكثر، فقد بلغ المتوسط الحسابي لمؤشر الدقة الإجمالية 91.5% Accuracy ومؤشر F1-Score 91.4%، في حين سجل مؤشر Precision نسبة 85.9%، ومؤشر الاستدعاء Recall نسبة 97.8%، مما يؤكد تحقيق نموذج شجرة القرار أداءا جيدا في مهمة التصنيف يتيح تعميم نتائجه بثقة تتراوح بين 85% و 91%. وعلاوة على ذلك، تراوحت قيم الانحراف المعياري لجميع هذه

المؤشرات بين 1% و3%، وهي معدلات انحراف منخفضة تعكس مدى تقارب النتائج بين طيات النموذج ،مما يستدل به إحصائياً على استقرار النموذج وموثوقيته العالية .

د- نسبة مساهمة المتغيرات في التنبؤ بالنموذج

بعد تطبيق نموذج شجرة القرار للتنبؤ بخروج الموظفين او بقائهم داخل الشركة وتقييم النموذج باستخدام كل من مصفوفة التحقق المتقاطع و المؤشرات العامة تم استخراج نسبة مساهمة كل متغير او الميزات من مختلف الميزات الموجودة في البيانات والتي تم اختيارها سابقا بناءا على ارتباطها القوي مع المتغير التابع وارتباطها الضعيف في ما بينها في بناء النموذج التنبؤي و ايضا تأثيرها الجماعي على دقة النموذج، حيث تهدف هذه النتائج إلى تحديد العوامل الأكثر تأثيرا على خروج الموظفين او بقائهم حيث أن هذه الخطوة تساعد بشكل كبير على معرفة أي من المتغيرات او الميزات التي يجب التركيز عليها والعمل على تحسينها أكثر بهدف ضمان بقاء بالموظفين و تجنب خروجهم

جدول: نسبة مساهمة المتغيرات في بناء نموذج شجرة القرار

نسبة المساهمة	المتغير
%9.99	DailyRate
%9.64	MonthlyIncome
%8.92	Age
%8.48	OverTime_Yes
%5.70	YearsInCurrentRole
%5.49	TrainingTimesLastYear
%4.95	YearsWithCurrManager
%4.74	PercentSalaryHike

المصدر: من إعداد الطالب

نلاحظ في الجدول أن أكبر نسبة مساهمة كانت في متغير DailyRate حيث قدرت نسبة مساهمته ب 9.99% و متغير MonthlyIncome الذي قدرت نسبة مساهمته ب9.64% حيث هذا يعني ان الجانب المادي و المالي مهم جدا و هذا ما يوضح ان الموظفين الذي يتقاضون

دخل مرتفع و ممتاز تكون احتمالية بقائهم و عدم خروجهم اعلى من الموظفين الذين يتقاضون راتب اقل وهذا ما يجعل المؤسسة تركز على جانب الرواتب من اجل الاحتفاظ اكثر بالموظفين ، يليه بعده المتغيرين Age و Overtime_Yes و اللذان نسبة مساهمتهما تقدران ب8.92% و 8.48% و هذا يعني ان العمر يلعب دور كبير في تحديد ما اذا كان الموظف سيخرج او لا و ايضا يعتبر متغير العمل الاضافي دورا هاما في تصنيف خروج الموظفين او بقائهم و هذا يدل على ضرورة بناء استراتيجية خاصة لكلاهما داخل المؤسسة من اجل الحفاظ على الموظفين، كما سجلت نسبة YearsInCurrentRole ب 5.70% وتدل هذه النسبة لتأثير سنوات البقاء في نفس المنصب للموظفين في الشركة أي أنه كلما كان عدد سنواتهم في نفس المنصب اكبر زاد ذلك من استقرارهم و يزيد هذا الاخير من احتمالية بقائهم داخل الشركة ، اما بقية المتغيرات TrainingTimesLastYear و YearsWithCurrManager و PercentSalaryHike فقد حصلا على نسبة مساهمة قدرت ب 5.49% و 4.95% و 4.74% و هذا يدل على انهم متغيرات مهمة ايضا بحيث يعتبر عدد مرات التدريب في السنة و عدد السنوات مع المدير و نسبة الزيادة في الراتب مؤثران ايضا على احتمالية خروج الموظفين او بقائهم داخل الشركة و هذا ما يتطلب ايضا استراتيجيات خاصة بهم ايضا.

ثانيا : الانحدار اللوجستي

هو اداة احصائية يمكننا من خلالها تقدير احتمالية وقوع الفئتين فئة الموظفين الذين سيخرجوا و فئة الموظفين سيبقوا ، ويمكن استيراد خوارزمية الانحدار اللوجستي في برمجية Python من خلال الأمر التالي :

الشكل: استيراد نموذج الانحدار اللوجستي

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression(random_state=42,C=5)
```

المصدر: من اعداد الطالب

أ- تقييم أداء النموذج باستخدام مصفوفة الالتباس

يمكن استخراج مصفوفة الالتباس لنموذج الانحدار اللوجستي برمجيا باستعمال الأمر التالي:

الشكل: أمر مصفوفة الالتباس الخاصة بنموذج الانحدار اللوجستي

```
confusion_matrix(y_test,y_pred_lr)
```

المصدر: من اعداد الطالب

حيث كانت النتائج التي اظهرتها مصفوفة الالتباس الخاصة بالانحدار اللوجستي و التي تحصلنا عليها من خلال ذلك الامر كالتالي:

	Actual Positive الموظف خرج فعلا	Actual Negative الموظف بقي فعلا
Predicted Positive توقع خروج الموظف	17 (TP)	41 (FP)
Predicted Negative توقع بقاء الموظف	14 (FN)	149 (TN)

المصدر: من إعداد الطالب

بناءا على مصفوفة الالتباس نلاحظ أن نموذج الانحدار اللوجستي قد توقع بدقة خروج 17 موظفاً للمؤسسة، وهي الحالات التي تمثل الإيجابية الحقيقية (True Positive) ،و أخطأ النموذج في تصنيف 41 موظفاً تنبا بخروجهم لكنهم بقوا، وهو ما يمثل الإيجابية الخاطئة (False Positive) ، كما تُشير المصفوفة إلى أن قيمة السلبية الخاطئة (False Negative) بلغت 14 حالة، لتمثل الموظفين الذين توقع النموذج بقاءهم في حين أنهم غادروا المؤسسة خرجوا و هذا يعني ان النموذج خطا اكثر مما اصاب في فئة الموظفين الذين سيخرجون ، و نجح النموذج في التصنيف الصحيح لـ 149 موظفاً تنبا ببقائهم وبقوا فيها فعلا، وهو ما يُعبر عن السلبية الحقيقي (True Negative) .

ب- مؤشرات الأداء العامة

جدول: المؤشرات العامة حسب نموذج الانحدار اللوجستي

Accuracy	0.7511
Precision	0.2931
Recall	0.5483
F1 Score	0.3820

المصدر: من إعداد الطالب

بعد بناء نموذج الانحدار اللوجستي على بيانات الموظفين تم تقييم أدائه باستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 75,11% وهي نسبة قوية تدل على الأداء ممتاز بشكل عام أي أن النموذج تمكن من تصنيف ما نسبته 81,90% من البيانات بشكل صحيح، لكن قدرت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة 29,31% وهي نسبة ضعيفة في التصنيف الفئة الايجابية، و بلغت نسبة مؤشر الاسترجاع Recall ب54,83% في ادراك الموظفين الذين سيخرجون وهي نسبة متوسطة وتعني أن النموذج استطاع الامساك 54,83% من الموظفين الذين سيخرجون لكن افلت منه نسبة كبيرة من الموظفين الذين سيخرجون قدرت نسبتهم ب45,17% حيث توقعها على أنها قيم سلبية إلا أنها كانت قيم ايجابية و ذلك راجع لأنه توقع بشكل صحيح في الفئة الثانية و اخطأ كثيرا في الفئة الاولى ، و قدرت نسبة مؤشر F1 Score ب38,20% وهي نسبة ضعيفة تدل على أن النموذج لم يحقق توازنا بين مؤشر الدقة الايجابية Precision ومؤشر الإسترجاع Recall..

ج- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Cross Validation

جدول: نتائج مصفوفة التحقق المتقاطع لنموذج الانحدار اللوجستي

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.7903	0.8116	0.82	0.7627
Fold2	0.7139	0.7554	0.7833	0.6558
Fold3	0.7341	0.7548	0.7428	0.7255
Fold4	0.7276	0.7418	0.7162	0.7395
Fold5	0.7255	0.7505	0.7414	0.7102
Mean cv	0.738	0.762	0.76	0.718

المصدر: من إعداد الطالب

نلاحظ من الجدول نتائج التحقق المتقاطع الخاص بنموذج الانحدار اللوجستي حيث نلاحظ نتائج النموذج اصبحت افضل بكثير مما هو عليه في التقسيمة الواحدة بعد ان عملنا علي التحقق المتقاطع و ذلك بسبب اختفاء التحيز في البيانات مما سمح للنموذج بالتعامل مع فئة الموظفين الذين خرجوا و الذين بقوا بشكل عادل و دقيق اكثر حيث بلغ المتوسط الحسابي لكل من المعايير F1-Scores الذي قيمته 73.8% و Precision الذي قيمته 76% و Recall الطي قيمته 71.8% ومعيار Accuracy قيمة 76.2% وهذا يشير إلى أن نموذج الانحدار اللوجستي يحقق أداءا جيدا في تصنيف الموظفين حيث يمكن بهذا تعميم النتئج بنسبة 71% إلى 76%، كما أن الانحراف المعياري لجميع المؤشرات تراوح بين 2% و الى 3%, هي درجات إنحراف منخفضة و تعكس مدى تقارب الطيات في النموذج وهذا يدل على استقرار النموذج و موثوقيته الكبيرة عبر مختلف الطيات الخمس .

د- نسبة مساهمة المتغيرات في التنبؤ بالنموذج

جدول: نسبة مساهمة المتغيرات في بناء نموذج الانحدار اللوجستي

نسبة المساهمة	المتغير
%13.37	OverTime_Yes
%10.97	MaritalStatus_Single
%7.97	Gender_Male
%6.94	JobRole_Laboratory Technician
%5.82	JobRole_Research Scientist
%5.78	EducationField_Life Sciences
%5.03	BusinessTravel_Travel_Frequently
%4.94	JobRole_Sales Executive

المصدر: من إعداد الطالب

نلاحظ في الجدول ان المتغير الاكثر مساهمة هو OverTime_Yes بنسبة 13.37% يليه المتغير MaritalStatus_Single بنسبة 10.97% و هذا يدل على ان الحالة الفردية للموظف تلعب دورا مهما

ايضا في المساهمة في احتمالية خروج الموظفين من المؤسسة و يليه بعده كل Gender_Male و JobRole_Laboratory Technician و JobRole_Research Scientist و JobRole_ResearchScientis و EducationField_Life_Sciences و BusinessTravel_Travel_Frequently و JobRole_Sales و Executive هذا يدل على ان العوامل الفردية للموظف و الخلفية الدراسية و القسم الذي يعمل فيه الموظف يلعبان دورا مهما ايضا في المساهمة في احتمالية خروج الموظف مما يتطلب وضع استراتيجية تخص هذه العوامل ايضا.

ثالثا : الغابة العشوائية

هي خوارزمية تعتمد على عدة اشجار قرار و تأخذ نتيجة أغلبية الاشجار التي تحقق اعلى دقة و يمكن استعمالها في تصنيف خروج الموظف او ويمكن استيرادها في برمجية Python حسب الأمر التالي :

الشكل: استيراد نموذج الغابة العشوائية

```
from sklearn.ensemble import RandomForestClassifier
rf=RandomForestClassifier(max_depth=20, n_estimators=200, random_state=42)
```

ا- تقييم أداء النموذج بإستخدام مصفوفة الإلتباس

يمكن استخراج مصفوفة الالتباس الخاصة بنموذج الغابة العشوائية برمجيا باستعمال الأمر التالي:

الشكل: أمر مصفوفة الالتباس الخاصة بنموذج الغابة العشوائية

```
confusion_matrix(y_test,y_pred_rf)
```

المصدر: من اعداد الطالب

حيث كانت النتائج التي اظهرتها مصفوفة الالتباس الخاصة بالغابة العشوائية و التي تحصلنا عليها كالتالي:

	Actual Positive الموظف خرج فعلا	Actual Negative الموظف بقي فعلا
Predicted Positive توقع خروج الموظف	7 (TP)	2 (FP)
Predicted Negative توقع بقاء الموظف	24 (FN)	188 (TN)

المصدر: من إعداد الطالب

بناءا على مصفوفة الالتباس نلاحظ أن نموذج الغابة العشوائية قد توقع بدقة خروج 7 موظفاً للمؤسسة، وهي الحالات التي تمثل الإيجابية الحقيقية (True Positive)،و أخطأ النموذج في تصنيف 2 موظفاً تنبا بخروجهم لكنهم بقوا، وهو ما يمثل الإيجابية الخاطئة (False Positive)، و تظهر المصفوفة أن قيمة السلبية الخاطئة (False Negative) بلغت 24 حالة، لتمثل الموظفين الذين توقع النموذج بقاءهم في حين أنهم خرجوا و هذا يعني ان النموذج خطا اكثر مما اصاب في فئة الموظفين الذين سيخرجون ، و نجح النموذج في التصنيف الصحيح لـ 188 موظفاً تنبا ببقائهم وبقوا فيها فعلا، وهو ما يُعبر عن السلبية الحقيقي (True Negative).

ب- مؤشرات الأداء العامة

جدول: المؤشرات العامة حسب نموذج الغابة العشوائية

Accuracy	0.8823
Precision	0.7777
Recall	0.2258
F1 Score	0.35

المصدر: من إعداد الطالب

بعد بناء نموذج الغابة العشوائية على بيانات الموظفين تم تقييم أدائه باستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 88,23% وهي نسبة قوية تدل على الأداء ممتاز بشكل عام أي أن النموذج تمكن من تصنيف ما نسبته 88,23% من البيانات بشكل صحيح، و قدرت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات

الإيجابية الخاطئة بنسبة 77,77% وهي نسبة جيدة لحد ما في التصنيف الفئة الايجابية، و بلغت نسبة مؤشر الاسترجاع Recall 22,58% في ادراك الموظفين الذين سيخرجون وهي نسبة ضعيفة وتعني أن النموذج استطاع الامساك ب25,58% من الموظفين الذين سيخرجون لكن افلت منه نسبة كبيرة من الموظفين الذين سيخرجون قدرت نسبتهم ب74,42% حيث توقعها على أنها قيم سلبية إلا أنها كانت قيم ايجابية و ذلك راجع لأنه توقع بشكل صحيح في الفئة الثانية و اخطأ كثيرا في الفئة الاولى ، و قدرت نسبة مؤشر F1 Score ب35% وهي نسبة ضعيفة تدل على أن النموذج لم يحقق توازنا بين مؤشر الدقة الايجابية Precision ومؤشر الإسترجاع Recall.

ج- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Cross Validation

جدول: نتائج مصفوفة التحقق المتقاطع لنموذج الغابة العشوائية

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.9767	0.9783	0.9767	0.9767
Fold2	0.9907	0.9913	0.9817	1
Fold3	0.9885	0.9891	0.9772	1
Fold4	0.9794	0.9804	0.9598	1
Fold5	0.9310	0.9392	0.9843	0.8831
Mean cv	0.973	0.975	0.976	0.971

المصدر: من إعداد الطالب

نلاحظ من الجدول نتائج التحقق المتقاطع الخاص بنموذج الغابة العشوائية حيث نلاحظ نتائج النموذج اصبحت افضل بكثير مما هو عليه في التقسيمة الواحدة بعد ان عملنا علي التحقق المتقاطع و ذلك بسبب اختفاء التحيز في البيانات مما سمح للنموذج بالتعامل مع فئة الموظفين الذين خرجوا و الذين بقوا بشكل عادل و دقيق اكثر حيث بلغ المتوسط الحسابي لكل من المعاييير F1-Scores الذي قيمته 97.3% و Precision الذي قيمته 97.6% و Recall الطي قيمته 97.1% ومعيار Accuracy قيمة 97.5% وهذا يشير إلى أن نموذج الغابة العشوائية يحقق أداءا جيدا في تصنيف الموظفين إذ أنه يمكن تعميم هذه النتائج بنسبة 97%، كما أن قيمة الإنحراف المعياري لجميع المؤشرات تتراوح بين 0.8% و الى 4%،هي درجات إنحراف منخفضة و تعكس مدى تقارب الطيات في النموذج وهذا يدل على إستقرار النموذج و موثوقيته الكبيرة عبر مختلف الطيات الخمس .

د- نسبة مساهمة المتغيرات في التنبؤ بالنموذج

جدول: نسبة مساهمة المتغيرات في بناء نموذج الغابة العشوائية

المتغير	نسبة المساهمة
MonthlyIncome	%9.98
Age	%8.46
OverTime_Yes	%7.66
DailyRate	%6.47
HourlyRate	%5.74
DistanceFromHome	%5.36
YearsWithCurrManager	%5.01
YearsInCurrentRole	%4.88

المصدر: من إعداد الطالب

نلاحظ في الجدول ان المتغير الاكثر مساهمة هو MonthlyIncome بنسبة 9.98% يليه المتغير Age بنسبة 8.46% و هذا يدل على تأثير عنصر العمر ايضا في المساهمة في احتمالية خروج الموظفين من المؤسسة و يليه بعده كل OverTime_Yes و DailyRate و HourlyRate و DistanceFromHome و YearsWithCurrManager و YearsInCurrentRole حيث كانت نسبة مساهمتهم متفاوتة و يدل هذا على اهميتهم ايضا.

رابعا : التدرج المعزز

هي خوارزمية تعتمد ايضا على اشجار القرار بحيث تعمل على اختبار شجرة قرار ضعيفة و تحسينها الى ان تصل الى اقصى دقة ممكنة و يمكن استعمالها في تصنيف خروج الموظف او ويمكن استيرادها في برمجية Python حسب الأمر التالي :

الشكل: استيراد نموذج التدرج المعزز

```
from sklearn.ensemble import GradientBoostingClassifier
xgb=GradientBoostingClassifier(n_estimators=200,random_state=42,max_depth=6,learning_rate=0.3)
```

المصدر: من اعداد الطالب

ا- تقييم أداء النموذج باستخدام مصفوفة الإلتباس

يمكن استخراج مصفوفة الالتباس الخاصة بنموذج التدرج المعزز باستعمال الأمر التالي:

الشكل: أمر مصفوفة الالتباس الخاصة بنموذج التدرج المعزز

```
confusion_matrix(y_test,y_pred_xgb)
```

المصدر: من اعداد الطالب

حيث كانت النتائج التي اظهرتها مصفوفة الالتباس الخاصة بالتدرج المعزز و التي تحصلنا عليها كالتالي:

	Actual Positive الموظف خرج فعلا	Actual Negative الموظف بقي فعلا
Predicted Positive توقع خروج الموظف	10 (TP)	2 (FP)
Predicted Negative توقع بقاء الموظف	21 (FN)	188 (TN)

المصدر: من إعداد الطالب

بناءا على مصفوفة الالتباس نلاحظ أن نموذج التدرج المعزز قد توقع بدقة خروج 10 موظفاً للمؤسسة، وهي الحالات التي تمثل الإيجابية الحقيقية (True Positive)،و أخطأ النموذج في تصنيف 2 موظفاً تنبا بخروجهم لكنهم بقوا، وهو ما يمثل الإيجابية الخاطئة (False Positive)، و تظهر المصفوفة أن قيمة السلبية الخاطئة (False Negative) بلغت 21 حالة، لتمثل الموظفين الذين توقع النموذج بقاءهم في حين أنهم خرجوا و هذا يعني ان النموذج خطا اكثر مما اصاب في فئة الموظفين الذين سيخرجون ، و نجح النموذج في التصنيف الصحيح لـ 188 موظفاً تنبا ببقائهم وبقوا فيها فعلا، وهو ما يُعبر عن السلبية الحقيقي (True Negative).

ب- مؤشرات الأداء العامة

جدول: المؤشرات العامة حسب نموذج التدرج المعزز

Accuracy	0.8959
Precision	0.8333
Recall	0.3225
F1 Score	0.4651

المصدر: من إعداد الطالب

بعد بناء نموذج التدرج المعزز على بيانات الموظفين تم تقييم أدائه باستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 89,59% وهي نسبة قوية تدل على الأداء ممتاز بشكل عام أي أن النموذج تمكن من تصنيف ما نسبته 89,59% من البيانات بشكل صحيح، و قدرت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة 83,33% وهي نسبة جيدة جدا ما في التصنيف الفئة الايجابية، و بلغت نسبة مؤشر الاسترجاع Recall ب32,25% في ادراك الموظفين الذين سيخرجون وهي نسبة ضعيفة وتعني أن النموذج استطاع الامساك ب32,25% من الموظفين الذين سيخرجون لكن افلت منه نسبة كبيرة من الموظفين الذين سيخرجون قدرت نسبتهم ب67,75% حيث توقعها على أنها قيم سلبية إلا أنها كانت قيم ايجابية و ذلك راجع لأنه توقع بشكل صحيح في الفئة الثانية و اخطأ كثيرا في الفئة الاولى ، و قدرت نسبة مؤشر F1 Score 46,51% وهي نسبة ضعيفة تدل على أن النموذج لم يحقق توازنا بين مؤشر الدقة الايجابية Precision ومؤشر الإسترجاع Recall.

ج- تقييم أداء النموذج بإستخدام مصفوفة التحقق المتقاطع Cross Validation

جدول: نتائج مصفوفة التحقق المتقاطع لنموذج التدرج المعزز

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.9701	0.9718	0.9590	0.9813
Fold2	0.9772	0.9783	0.9555	1
Fold3	0.9817	0.9826	0.9641	1
Fold4	0.9706	0.9718	0.9429	1
Fold5	0.9297	0.9370	0.9648	0.8971
Mean cv	0.965	0.968	0.957	0.975

المصدر: من إعداد الطالب

نلاحظ من الجدول نتائج التحقق المتقاطع الخاص بنموذج التدرج المعزز حيث نلاحظ نتائج النموذج اصبحت افضل بكثير مما هو عليه في التقسيمة الواحدة بعد ان عملنا علي التحقق المتقاطع و ذلك بسبب اختفاء التحيز في البيانات مما سمح للنموذج بالتعامل مع فئة الموظفين الذين خرجوا و الذين بقوا بشكل عادل و دقيق اكثر حيث بلغ المتوسط الحسابي لكل من المعايير F1-Scores الذي قيمته 96.5% و Precision الذي قيمته 95.7% و Recall الطي قيمته 97.5% ومعيار Accuracy قيمة 96.8% وهذا يشير إلى أن نموذج التدرج المعزز يحقق جيدا في تصنيف الموظفين إذ أنه يمكن تعميم هذه النتائج بنسبة بين 95% و97%، كما أن قيمة الإنحراف المعياري لجميع المؤشرات تتراوح بين 0.7% و الى 3%, هي درجات إنحراف منخفضة و تعكس مدى تقارب الطيات في النموذج وهذا يدل على إستقرار النموذج و موثوقيته الكبيرة عبر مختلف الطيات الخمس .

د- نسبة مساهمة المتغيرات في التنبؤ بالنموذج

جدول: نسبة مساهمة المتغيرات في بناء نموذج التدرج المعزز

المتغير	نسبة المساهمة
MonthlyIncome	%12.77
OverTime_Yes	%11.90
Age	%9.58
DailyRate	%8.64
YearsWithCurrManager	%5.40
MaritalStatus_Single	%5.27
YearsInCurrentRole	%4.60
DistanceFromHome	%4.36

المصدر: من إعداد الطالب

نلاحظ في الجدول ان المتغير الاكثر مساهمة هو MonthlyIncome بنسبة 12.77% يليه المتغير OverTime_Yes بنسبة 11.90% و هذا يدل على ان هاذين العنصرين مهمان جدا كما قلنا سابقا في المساهمة في احتمالية خروج الموظفين او بقائهم و يليه بعده كل Age و DailyRate و YearsWithCurrManager و MaritalStatus_Single و YearsInCurrentRole و DistanceFromHome حيث كانت نسبة مساهمتهم متفاوتة و يدل هذا على اهميتهم ايضا.

خامسا : الة الدعم الشعاعي

هو نموذج يعتمد على رسم خط يفصل بين النقاط من اجل الحصول على تصنيف لفئتين و يمكن استعمالها في تصنيف خروج الموظف او ويمكن إستيرادها في برمجية Python من خلال الامر الاتي:

الشكل: إستيراد نموذج الة الدعم الشعاعي

```
from sklearn.svm import  SVC
svm=SVC(kernel='rbf',probability=True,gamma=0.94)
```

المصدر: من اعداد الطالب

ا- تقييم أداء النموذج باستخدام مصفوفة الالتباس

يمكننا استخراج مصفوفة الالتباس الخاصة بنموذج الة الدعم الشعاعي برمجيا باستعمال الأمر التالي:

الشكل: أمر مصفوفة الالتباس الخاصة بنموذج الة الدعم الشعاعي

```
confusion_matrix(y_test,y_pred_svm)
```

المصدر: من اعداد الطالب

حيث كانت النتائج التي اظهرتها مصفوفة الالتباس الخاصة بالة الدعم الشعاعي و التي تحصلنا كالتالي:

	Actual Positive الموظف خرج فعلا	Actual Negative الموظف بقي فعلا
Predicted Positive توقع خروج الموظف	5 (TP)	0 (FP)
Predicted Negative توقع بقاء الموظف	26 (FN)	190 (TN)

المصدر: من إعداد الطالب

تظهر مؤشرات مصفوفة الالتباس أن نموذج الة الدعم الشعاعي قد توقع بدقة خروج 5 موظفاً للمؤسسة، وهي الحالات التي تمثل الإيجابية الحقيقية (True Positive)،و أخطأ النموذج في تصنيف 0 موظفاً تنبا بخروجهم لكنهم بقوا، وهو ما يمثل الإيجابية الخاطئة (False Positive) و هذا يدل على ان نموذج الة الدعم الشعاعي قوي جدا في تصنيف فئة الموظفين الباقين في المؤسسة ، و تظهر المصفوفة أن قيمة السلبية الخاطئة (False Negative) بلغت 26 حالة، لتمثل الموظفين الذين توقع النموذج بقائهم في حين أنهم خرجوا و هذا يعني ان النموذج خطا اكثر مما اصاب في فئة الموظفين الذين سيخرجون ، و نجح النموذج في التصنيف الصحيح لـ 190 موظفاً تنبا ببقائهم وبقوا فيها فعلا، وهو ما يمثل عن السلبية الحقيقي (True Negative).

ب- مؤشرات الأداء العامة

جدول: المؤشرات العامة حسب نموذج الة الدعم الشعاعي

Accuracy	0.8868
Precision	1.0
Recall	0.1612
F1 Score	0.2777

المصدر: من إعداد الطالب

بعد بناء نموذج الة الدعم الشعاعي على بيانات الموظفين تم تقييم أدائه باستخدام المؤشرات السابقة يتبين أن دقة النموذج العامة Accuracy بلغت نسبة 88,68% وهي نسبة قوية تدل على الأداء ممتاز بشكل عام أي أن النموذج تمكن من تصنيف ما نسبته 88,68% من البيانات بشكل صحيح، و قدرت الدقة الإيجابية للنموذج Precision والتي تعكس قدرة النموذج على تقليل قدرة التوقعات الإيجابية الخاطئة بنسبة 100% وهي نسبة مثالية ما في التصنيف الفئة الايجابية، و بلغت نسبة مؤشر الاسترجاع Recall ب16,12% في ادراك الموظفين الذين سيخرجون وهي نسبة ضعيفة وتعني أن النموذج استطاع الامساك ب16,12% من الموظفين الذين سيخرجون لكن افلت منه نسبة كبيرة من الموظفين الذين سيخرجون قدرت نسبتهم ب83,88% حيث توقعها على أنها قيم سلبية إلا أنها كانت قيم ايجابية و ذلك راجع لأنه توقع بشكل صحيح في الفئة الثانية و اخطأ كثيرا في الفئة الاولى ، و قدرت نسبة مؤشر F1 Score ب27,77% وهي نسبة ضعيفة تدل على أن النموذج لم يحقق توازنا بين مؤشر الدقة الايجابية Precision ومؤشر الإسترجاع Recall.

ج- تقييم أداء النموذج باستخدام مصفوفة التحقق المتقاطع Cross Validation

جدول: نتائج مصفوفة التحقق المتقاطع لنموذج الة الدعم الشعاعي

Folds	F1-Scores	Accuracy	Precision	Recall
Fold1	0.9766	0.9783	0.9812	0.9720
Fold2	0.9862	0.9870	0.9728	1
Fold3	0.9794	0.9804	0.9639	0.9953
Fold4	0.9793	0.9804	0.9681	0.9906
Fold5	0.9211	0.9305	0.9739	0.8738
Mean cv	0.968	0.971	0.972	0.966

المصدر: من إعداد الطالب

نلاحظ من الجدول نتائج التحقق المتقاطع الخاص بنموذج الة الدعم الشعاعي حيث نلاحظ نتائج النموذج اصبحت افضل بكثير مما هو عليه في التقسيمة الواحدة بعد ان عملنا علي التحقق المتقاطع و ذلك بسبب اختفاء التحيز في البيانات مما سمح للنموذج بالتعامل مع فئة الموظفين الذين خرجوا و الذين بقوا بشكل عادل و دقيق اكثر حيث بلغ المتوسط الحسابي لكل من المعايير F1-Scores الذي قيمته 96.8% و Precision الذي قيمته 97.2% و Recall الطي قيمته 96.6% ومعيار Accuracy قيمة 97.1% وهذا يشير إلى أن نموذج الة الدعم الشعاعي يحقق أداءا جيدا في تصنيف الموظفين إذ أنه يمكن تعميم هذه النتائج بنسبة بين 96% و97%، كما أن قيمة الإنحراف المعياري لجميع المؤشرات تتراوح بين 0.5% و الى 5%, هي درجات إنحراف منخفضة و تعكس مدى تقارب الطيات في النموذج وهذا يدل على إستقرار النموذج و موثوقيته الكبيرة عبر مختلف الطيات الخمس .

د- نسبة مساهمة المتغيرات في التنبؤ بالنموذج

جدول: نسبة مساهمة المتغيرات في بناء نموذج الة الدعم الشعاعي

المتغير	نسبة المساهمة
MonthlyIncome	%12.77
OverTime_Yes	%11.90
Age	%9.58
DailyRate	%8.64
YearsWithCurrManager	%5.40
WorkLifeBalance	%5.27
YearsInCurrentRole	%4.60
DistanceFromHome	%4.36

المصدر: من إعداد الطالب

نلاحظ في الجدول ان المتغير الاكثر مساهمة هو MonthlyIncome بنسبة 9.98% يليه المتغير OverTime_Yes بنسبة 11.90% و يليه بعده كل Age و DailyRate و YearsWithCurrManager و WorkLifeBalance و YearsInCurrentRole و DistanceFromHome حيث كانت نسبة مساهمتهم متفاوتة و يدل هذا على اهميتهم ايضا.

المطلب الثالث: المقارنة بين الدراسة الحالية و الدراسة السابقة

سنقوم في هذا المطلب بالمقارنة بين دراستنا الحالية و بين دراسة سابقة عملت على نفس قاعدة البيانات و ذلك من اجل اثبات موثوقية نتائجنا و صحتها .

النموذج	الدراسة الحالية	الدراسة السابقة
شجرة القرار	%91.5	%80.5
الانحدار اللوجستي	%76.2	–
الغابة العشوائية	%97.5	%87.3
التدرج المعزز	%96.8	%87.3
الة الدعم الشعاعي	%97.1	%88.6
النموذج الهجين	–	%88.7

المصدر: من اعداد الطالب

نلاحظ من خلال الجدول ان افضل نموذج حقق اعلى دقة في الدراسة الحالية الخاصة بنا كان الغابة العشوائية بنسبة دقة 97,5% بينما كان افضل نموذج حقق اعلى دقة في الدراسة السابقة كان النموذج الهجين بنسبة دقة 88,7% حيث نرى ان النتائج التي حققناها في هذه الدراسة افضل من النتائج التي حققتها الدراسة السابقة و ذلك بسبب الاختيار الافضل للميزات التي ترفع من الدقة و ايضا من خلال ضبط اعدادات النماذج للوصول لأعلى دقة تحققها و ايضا استخدام التحقق المتقاطع و الذي قلل من تحيز البيانات على عكس الدراسة السابقة التي اعتمدت على نتائج تقسيمة واحدة و هذا ما يثبت ان نماذج دراستنا الحالية كان ادائها ممتاز و موثوق بنسبة اعلى من الدراسة السابقة.

خلاصة الفصل:

ركز هذا الفصل على الجانب التطبيقي من خلال استغلال بيانات شركة IBM المستمدة من منصة (Kaggle)، بغرض تطوير نموذج تنبؤي لتصنيف الموظفين وتوقع قرار مغادرتهم أو بقائهم في المؤسسة. وقد شمل العمل معالجة البيانات وتحليلها بعمق لاستخلاص الميزات الأكثر تأثيراً في بناء النموذج، ومن ثم توظيف مجموعة من خوارزميات التعلم الآلي المتخصصة في التنبؤ لتحقيق أهداف الدراسة. كما تم تقييم أداء هذه النماذج بناءً على مؤشرات دقة متعددة لتحديد النموذج الأمثل. واختتم الفصل بإجراء مقارنة بين مخرجات الدراسة الحالية ودراسة سابقة اعتمدت على نفس قاعدة البيانات والنماذج، حيث أظهرت النتائج تفوق الدراسة الحالية من حيث الدقة العامة.

الخاتمة

الخاتمة

في الختام، يُعد التنبؤ بمغادرة الموظفين (Attrition) ركيزة أساسية تدعم المؤسسات الاقتصادية في اتخاذ قرارات استراتيجية للاحتفاظ بكفاءاتها، وذلك من خلال التحليل العميق لبيانات الموظفين وخصائصهم الوظيفية والشخصية. وقد أثبتت هذه الدراسة الفعالية العالية لتقنيات تعلم الآلة في التعامل مع بيانات الموارد البشرية، نظراً لقدرتها الفائقة على معالجة المتغيرات المتعددة والمعقدة المرتبطة بالعنصر البشري.لقد نجحت الدراسة في بناء نموذج تنبؤي قادر على تصنيف الموظفين بدقة عالية، مما يتيح استشراف قراراتهم بالمغادرة بشكل استباقي. ومن خلال التجربة والمقارنة بين عدة نماذج تصنيفية، برز نموذج الغابة العشوائية كأفضل النماذج من حيث الدقة التنبؤية العامة. و اكدت الدراسة ضرورة تبني المؤسسات الحديثة لتقنيات تعلم الآلة حيث لم يعد اتخاذ القرار يقتصر على تحليل الماضي أو الحاضر، بل اصبح يرتكز بشكل جوهري على التنبؤ بالأحداث المستقبلية لضمان الاستدامة والاستقرار داخل المؤسسات الاقتصادية.

اختبار الفرضيات:

– تبين في هذه الدراسة ان نموذج التدرج المعزز لا يعتبر افضل نموذج للتنبؤ بخروج او مغادرة الموظفين بل توصلت الدراسة إلى أن أفضل نموذج من حيث الدقة العامة (Accuracy) هو نموذج الغابة العشوائية (Random Forest)، حيث حقق نسبة دقة بلغت 97.5% من بين جميع النماذج التي تم تجريبها و هذا يثبت ان الفرضية خاطئة.

– استخدام تقنية التحقق المتقاطع ساهم في تقليل العشوائية في تقسيم البيانات و جعل النماذج تتدرب و تختبر على كل البيانات مما جعلها افضل دقة و موثوقية و قابلة للتعميم على بيانات جديدة و هذا اثبت على صحة الفرضية.

نتائج الدراسة:

– تبين من خلال رؤية مساهمة الميزات او المتغيرات ان الراتب و عدد ساعات العمل يؤثران على قرار الموظف في المغادرة او البقاء في المؤسسة.

– يعتبر كل من نموذج التدرج المعزز و الة الدعم الشعاعي و الغابة العشوائية من اهم النماذج للتنبؤ بسلوك الموظفين داخل المؤسسة.

– أثبتت خطوة موازنة البيانات الغير متوازنة انها مهمة جدا لتجنب انحياز النموذج نحو الفئة الاكثرية مما يعطي نتائج اكثر دقة و موثوقية و اقل انحيازا.

توصيات الدراسة:

– يجب على المؤسسات ان تضع ساعات عمل مناسبة و رواتب مناسبة و التركيز على الموظفين الاقل سنا بحيث تساهم هذه الاجراءات في الحد من مغادرة الموظفين و زيادة نسبة بقائهم في المؤسسة.

– يفضل استخدام تقنيات اختيار الميزات و المتغيرات اليا للوصول لأفضل مجموعة من المتغيرات التي تحقق اعلى دقة تنبؤية و ايضا رؤية مدى مساهمة المتغيرات في نماذج التنبؤ من اجل تفسير النتاج و اتخاذ القرارات المناسبة بناءا عليها.

– يجب على المؤسسات الاقتصادية ان تعتمد على التحليل التنبؤي بدل الاكتفاء بالتحليل الوصفي في هذا العصر

قائمة المراجع

المراجع العربية:

1- أسعد أحمد محمد عكاشة (2008) اثر الثقافة التنظيمية على مستوى الأداء الوظيفي، دراسة تطبيقية على شركة الاتصالات بفلسطين .رسالة ماجستير في إدارة الأعمال، غزة.

2- بلعرابي عبد الكريم (جانفي 2017). الحالات القانونية لإنهاء خدمة الموظف العمومي في القانون الجزائري .دفاتر السياسة والقانون، كلية الحقوق، جامعة ورقلة، العدد 16.

3- بوغازي فريدة (2015). فعالية تطبيق تقنيات التنبؤ بالمبيعات في المؤسسة: دراسة تطبيقية بمؤسسة / GNL سكيكدة بالجزائر .مجلة الباحث الاقتصادي، العدد 04.

4- حنان بن عوالي. (2013). التنبؤ بالطلب كمؤثر أساسي في عملية تخطيط الإنتاج .مجلة الباحث الإقتصادي، العدد 01.

5- رشيد خطارة (2019) الذكاء الانفعالي وعلاقته بالالتزام التنظيمي .أطروحة دكتوراه علم النفس التربوي، ورقلة، الجزائر.

6- السويدان، عبد الرحمن (2015) علم النفس النمو: الطفولة والمراهقة .دار الفكر.

7- الشرقاوي، عبد الرحمن (2019) إدارة الموارد البشرية: مدخل استراتيجي .القاهرة: دار الفكر العربي.

8- الزيات، أحمد (2018) علم النفس التربوي وتطبيقاته في التعليم .القاهرة: دار الفكر العربي.

9- عادل حسن. (1969) الكفاية الإنتاجية للأفراد في الصناعة .دار الجامعات المصرية، الإسكندرية.

10- فاروق عبد الفتاح موسى (1985) أسس السلوك الإنساني .دار عالم الكتاب.

11- فضيلة فراحي. (2022) استخلاص المتغيرات الاقتصادية المحفزة لنشاط بورصة الجزائر باستخدام خوارزمية متجه الدعم الآلي .مجلةLes Cahiers du Cread ، المجلد 38.

12- كردودي سهام، وبن قدور علي التنبؤ بالشبكات العصبية الاصطناعية كدعامة للمراجعة التحليلية في عملية التدقيق: حالة مركب تكرير الملح SEL.NA.E (بسكرة) في الفترة 2010-2014 .مجلة الإستراتيجية والتنمية.

13- ليزا تاغليفيري. (2022) كتاب مشاريع تعلم الآلة بايثون.

14- محمد أمين دعيش، ومحمد ساري. (2017). نموذج الانحدار اللوجستي: مفهومه، خصائصه، تطبيقاته مع مثال تطبيقي الانحدار اللوجستي ثنائي الحدين على برنامج spss مجلة السراج في التربية وقضايا المجتمع، العدد الأول.

15- محمد لحلح. (2020) *كتاب مدخل إلى الذكاء الاصطناعي وتعلم الآلة*.

16- مولاي علي الزهر. (2017). مقال السلوك التنظيمي بين النظرية والتطبيق .*مجلة الآداب والعلوم الاجتماعية*، العدد 17.

17- ميلاد وزان *كتاب تعلم الآلة وعلم البيانات*.

ثانياً: المراجع باللغة الاجنبية

1- Cihat Polat. (2008). Forecasting as a Strategic Decision Making Tool: A Review and Discussion with Emphasis on Marketing Management. European Journal of Scientific Research.

2- Craig Starbuck. (2023). The Fundamentals of People Analytics. Springer.

3- Deepti Sharma, Sellvasundaram, Prasanta Chatterjee Biswas, Suresh Kumar, Ganesh, and Kinjal Parmar. (2024). Machine Learning and HRM: A Path to Efficient Workforce Management. Journal of Informatics Education and Research.

4- Dkhaz Mustafa Abdullah. (2021). Machine Learning Applications based on SVM Classification: A Review. Qubahan Academic Journal.

5- Eteng, Joy Umuolo, Bernard Samuel Eventus, Chenui, Armand Phobuh, Bassey Bassey Ubi, Okon, Utibe Essien & Etim, Umo Ephraim. (2025). Employee Attrition and Organisational Sustainability. International Journal of Social Sciences and Management Research.

7- Hassan Ahmed Salman, Ali Kalakesh, and Amani Steiti. (2024). Random Forest Algorithm Overview. Babylonian Journal of Machine Learning.

8- Ibomoiye Domor Mienie and Norbert Jere. (2024). A Survey of Decision Trees: Concepts, Algorithms, and Applications. IEEE Access.

9- Inemesit Godwin Enyina, Samuel Okurebia and Kingsley Uwa. (2023). Employee's Work Behaviour: Absenteeism, Embezzlement, & Organizational Performance. International Journal of Business and Management Review.

10- Iqbal H. Sarker. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science.

11- Khirod Chandra Panda and Shobhit Agrawal. (2021). Predictive Analytics: An Overview of Evolving Trends and Methodologies. Journal of Scientific and Engineering Research.

12- Maximilian Zellner, Ali Abbas, David Budescu, and Aram Galstyan. (2021). A survey of human judgement and quantitative forecasting methods. Royal Society Open Science.

13- Mazin Aledeinat, Zeyad Alkhazali, Hakem Sharari, Malek Alsoud. (2021). TOWARDS A NEW MODEL FOR WORK-LIFE BALANCE: A FACTORS-REVIEW APPROACH. Journal of Management Information and Decision Sciences.

14- ME, Sermpene et Al. (2002). Job Satisfaction in Relation to Organizational Culture. Journal of Industrial Psychology, 28(2).

15- Nugraha, Putri, Mubarok, Hadani. (2024). The Effect of Salary on Employee's Performance: A Study in Central Java. Proceedings of the 7th Global Conference on Business, Management, and Entrepreneurship (GCBME 2022).

16- Pritika Talwar, Shubham, and Komalpreet Kaur. (2024). EXPLORING CLUSTERING TECHNIQUES IN MACHINE LEARNING. International Journal of Creative Research Thoughts.

17- Régis Bourbonnais and Jean-Claude Usunier. (2008). Sales Forecasting: Between Theory and Practice.

18- Rob J. Hyndman and George Athanasopoulos. (2014). Forecasting: Principles and Practice.

19- Rodrigo Miranda Cabrera, Manuel Leon Ganchozo, Rodríguez Salazar Oswaldo Rene, Rafael Wilfredo Rojas Bujaico, Héctor Huamán Samaniego, John Fredy Rojas Bujaico. (2022). Impact Of Machine Learning In Human Resource Management: Towards The Modernization Of Leadership. Journal of Positive School Psychology.

20- Sebastian Raschka and Vahid Mirjalili. (2019). Python Machine Learning.

21- Sonal Vinayak Kasari and Yash Arun Mali. (2023). A Comparative Analysis of Logistic Regression and Decision Tree Algorithms for Fake News Detection. International Journal of Novel Research and Development.

22- Venkata Kalyan Chakravarthy Mandavilli. (2025). Harnessing the Power of Predictive Analytics: Transforming Business Intelligence. European Journal of Computer Science and Information Technology.

23- Yu Shi. (2023). Application of Improved Linear Regression Algorithm in business Behavior Analysis. *Elsevier*.

24- Yue Liu, Shulei Wu, Zhongqiang Wu, Shuangshuang Zhou. (2024). Application of gradient boosting machine in satellite-derived bathymetry using Sentinel-2 data for accurate water depth estimation in coastal environments. *Elsevier*.

ثالثاً: المواقع الإلكترونية ومصادر البيانات

1- https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

2- https://www.geeksforgeeks.org

ملخص الدراسة

يُعدّ سلوك الموظفين عنصرًا مهمًا في فهم العلاقة بين الموظفين والمؤسسات وتعزيزها، ومن بين السلوكيات التي تمثل خطرًا وتهديدًا على المؤسسات سلوك مغادرة الموظفين، إذ يساعد التنبؤ به مسبقًا على الحد من معدل المغادرة، وزيادة ولاء الموظفين، واتخاذ قرارات استباقية مبنية على أسس علمية ونتائج دقيقة، إضافة إلى حماية المؤسسات من ارتفاع التكاليف وخسارة الكفاءات البشرية.

تهدف هذه الدراسة إلى بناء نموذج قادر على التنبؤ بسلوك الموظفين من خلال تصنيفهم إلى فئتين: موظفون سيغادرون المؤسسة وموظفون سيستمرون فيها، وذلك بالاعتماد على تقنيات تعلم الآلة. وقد تم تطبيق المنهج الوصفي التحليلي، حيث اعتمدت الدراسة على قاعدة بيانات أنشأتها شركة IBM وتم الحصول عليها من منصة Kaggle.

تم اختبار خمسة نماذج لتعلم الآلة، وهي: شجرة القرار(Decision Tree) ، والتدرج المعزز (Gradient Boosting)، والانحدار اللوجستي(Logistic Regression) ، والغابات العشوائية (Random Forest)، وآلة الدعم الشعاعي(SVM) ، مع إجراء مقارنة بينها باستخدام مؤشرات التقييم العامة المتمثلة في: الدقة(Accuracy) ، والاستدعاء(Recall) ، والإحكام(Precision) ، ودرجة(F1-Score) ، وذلك بالاعتماد على أسلوب التحقق المتقاطع بهدف اختيار النموذج الأكثر كفاءة في التنبؤ.

وقد توصلت الدراسة إلى أن أفضل نموذج من حيث الدقة العامة (Accuracy) هو نموذج الغابة العشوائية(Random Forest) ، حيث حقق نسبة دقة بلغت 97.5%.

الكلمات المفتاحية :سلوك الموظفين، تعلم الآلة، التنبؤ، مغادرة الموظفين.

Abstract

Employee behavior is considered an important factor in understanding and strengthening the relationship between employees and organizations. Among the behaviors that pose a risk and threat to organizations is employee turnover behavior. Predicting this behavior in advance helps organizations reduce employee turnover, increase employee loyalty, make proactive decisions based on scientific foundations and accurate results, and protect themselves from rising costs and the loss of human talent.

This study aims to build a model capable of predicting employee behavior by classifying employees into two categories: those who are likely to leave the organization and those who are likely to stay, using machine learning techniques. The study adopted the descriptive analytical approach and relied on a dataset created by IBM and obtained from the Kaggle platform.

Five machine learning models were tested, namely: Decision Tree, Gradient Boosting, Logistic Regression, Random Forest, and Support Vector Machine (SVM). A comparison was conducted using general evaluation metrics, including Accuracy, Recall, Precision, and F1-Score, through cross-validation in order to determine the most efficient predictive model.

The findings of the study revealed that the Random Forest model achieved the best overall performance in terms of accuracy, reaching 97.5%.

Keywords: Employee Behavior, Machine Learning, Prediction, Employee Attrition.