ملخص:

يعد مشروع الفصل الخاص بك فرصة لك لاستكشاف مشكلة تعلم الآلة المثيرة للاهتمام من اختيارك في سياق مجموعة بيانات حقيقية. يجب أن يكون الغرض من المشروع مرتبطًا ببعض جوانب المادة ولكن قد يستكشف طريقًا لم تتم معالجته في الفصل. يمكن تنفيذ المشاريع بشكل فردي أو في فرق مكونة من طالبين. سيكون مشروعك بقيمة 20% من درجتك النهائية في الفصل الدراسي، وسيتضمن 4 مخرجات:

1. الاقتراح: (4 فبراير) على الأكثر صفحة واحدة، بخط 12 نقطة، وتباعد فردي، وهوامش 1 بوصة (10%)

2. تقرير منتصف الطريق: (25 فبراير) 4-5 صفحات (25%)

3. عرض الملصق: (21 مارس) (20%)

4. التقرير النهائي: (21 مارس) (45%)

مقترح المشروع (تاريخ الاستحقاق: الأحد، 4، الساعة 23:59 مساءً) (10%)

من أجل المساعدة في توجيه اختيارك للمشروع، يتعين عليك تقديم مقترح موجز (على الأكثر صفحة واحدة، خط 12 نقطة، تباعد فردي، هوامش 1 بوصة) يصف فكرة المشروع. يجب أن يكون الاقتراح صفحة واحدة كحد أقصى. يشمل المعلومات التالية:

• عنوان المشروع

فكرة المشروع. وينبغي أن يكون هذا ما يقرب من فقرتين. يجب أن تحدد نوع المشروع، والمشكلة التي تخطط لمعالجتها، والدافع الذي يجعلك تجد المشكلة مهمة أو مثيرة للاهتمام، وأي عمل سابق تعرفه بالفعل (إن أمكن)، ونهج مؤقت تقريبي لحل المشكلة.

• مجموعة البيانات. قم بوصف مجموعة البيانات، مثل عدد الأعمدة والصفوف ومصدر مجموعة البيانات وما إلى ذلك.

• البرامج التي سوف تحتاج إلى كتابتها.

زميل الفريق (إن وجد) وقسم العمل. أتوقع أن تكون المشاريع التي يتم تنفيذها في مجموعة أكثر أهمية من المشاريع التي يتم تنفيذها بشكل فردي.

• إنجاز منتصف المدة: ما الذي ستكمله بحلول 25 فبراير؟ ومن المتوقع هنا الحصول على نتائج تجريبية من نوع ما.

حد الصفحة: صفحة واحدة

ملحوظة: إذا كنت تواجه صعوبة في كتابة المقترح فلا تتردد في التواصل مع المدرب (د. علي أبوراس).

تقرير منتصف الطريق (25 فبراير) 4-5 صفحات (25%)

ينبغي أن يكون هذا التقرير قصيرًا من 2 إلى 4 صفحات، ويكون بمثابة نقطة تفتيش. يجب أن يتكون من نفس أقسام تقريرك النهائي (الخلفية، الطريقة، التجربة، الاستنتاج، والمراجع). يجب أن تكون المقدمة وأقسام العمل ذات الصلة في شكلها النهائي تقريبًا؛ يجب أن يكون القسم الخاص بالطريقة المقترحة قد انتهى تقريبًا؛ ستحتوي الأقسام الخاصة بالتجارب والاستنتاجات على النتائج التي حصلت عليها، بالإضافة إلى أماكن للنتائج الإضافية التي تخطط/تأمل في الحصول عليها.

نظام الدرجات لتقرير المشروع:

• 70% للطريقة المقترحة ونتائج بعض التجارب حتى الآن 30% لتصميم التجارب القادمة

عرض الملصق (21 مارس) (20%)

يجب على جميع أعضاء المشروع الحضور خلال ساعات العرض. ستكون الجلسة مفتوحة للجميع (إن أمكن).

يمكنك إنشاء مجموعة من شرائح العرض التقديمي العادية، وطباعة كل واحدة منها على قطعة من الورق (بحجم letter)، ووضعها جميعًا معًا على لوحة الملصقات. سأقدم لك القالب إذا كنت في حاجة إليه.

التقرير النهائي: (21 مارس) (45%)

يجب أن يتضمن التقرير النهائي حوالي أربع (3-4) صفحات من النص لكل شخص (لا يشمل ذلك

الأرقام) بنفس تنسيق الاقتراح وتقرير منتصف الطريق. يتضمن التقرير النهائي أقسامًا مثل (الخلفية، الطريقة، التجربة، الاستنتاج، والمراجع).

نوع المشروع/المشاريع المقترحة

هناك أنواع مختلفة من المشاريع التي يمكنك وضعها في الاعتبار:

1. تطبيق التقنيات التي تعلمتها: قد يكون المشروع عمليًا جدًا من حيث تطبيق التقنيات التي تعلمتها في الدورة على مشكلة حقيقية مثل تصنيف رسائل البريد الإلكتروني.

2. قد يتضمن المشروع تصميم أو تكييف الخوارزميات الموجودة مع فئة جديدة من المشاكل. على سبيل المثال، كيف يمكننا حل مهام التصنيف المتعددة ذات الصلة؟

3. مقارنة الخوارزميات: طوال الدورة، كنا نناقش الخوارزميات المختلفة وخصائصها. في كثير من الأحيان، لا تعمل الخوارزميات كما هو متوقع وقد تحتاج الخوارزميات إلى تكييفها أو تعديلها لتتناسب بشكل أفضل مع الافتراضات الكامنة في البيانات. ما العمل الذي يجب القيام به لتكييف النموذج مع مجموعة مثيرة للاهتمام من البيانات التي عثرت عليها؟ كيف تعمل الخوارزميات المختلفة على نفس مجموعة البيانات؟

4. المعلومات المفقودة: تتضمن مشكلات التصنيف المختلفة في العالم الحقيقي مكونات مفقودة في متجهات الإدخال. كيف يمكنك التعامل مع هذه المعلومات المفقودة؟ هل تتوقع أن تتدهور طريقتك بسرعة في حالة فقدان المزيد من المعلومات؟

5. تحديد المعلمات الفائقة (على سبيل المثال، وظيفة kernel في SVMs): تقوم خوارزميات التعلم الآلي تلقائيًا بضبط (تعلم) معلماتها الداخلية بناءً على البيانات. ومع ذلك، توجد مجموعة فرعية من المعلمات التي لم يتم تعلمها ويجب تكوينها يدويًا. يمكن أن يعتمد أداء النموذج على اختيار المعلمات الفائقة الخاصة به. أربع إستراتيجيات تحسين شائعة الاستخدام: بحث الشبكة، والبحث العشوائي، وتسلق التلال، والتحسين الافتراضي.

6. طرق حل الطبقات غير المتوازنة. تشير البيانات غير المتوازنة عادةً إلى مشكلة تتعلق بمشاكل التصنيف حيث لا يتم تمثيل الفئات بشكل متساوٍ. هناك خوارزميات منهجية يمكنك استخدامها لإنشاء عينات تركيبية. إحدى الطرق البسيطة لإنشاء عينات تركيبية هي أخذ عينات عشوائية من السمات من مثيلات في فئة الأقلية. وهو يتألف من إزالة عينات من فئة الأغلبية (نقص العينات) و/أو إضافة المزيد من الأمثلة من فئة الأقلية (الإفراط في أخذ العينات). أشهر هذه الخوارزميات تسمى SMOTE أو تقنية الإفراط في أخذ العينات للأقليات الاصطناعية.

ملحوظة: لا داعي للقلق بشأن الحصول على نتائج رائعة. تعد الفكرة وفهمك لقضايا التعلم الآلي المعنية أكثر أهمية بكثير من الحصول على نتائج رائعة.