

1- Post 1

فى الطبعى كنت بروج اعمل label encoder لاي categorical attribute وساعات كنت اعمل one_hot encoder من غير ما اعرف ده امتا وده امتا بس اتضح انى ده مش صح لان ال machine learning model ممكن يلاحظ فى حالة زى ال label encoder or Ordinal Encoder انى فى pattern معين زى ترتيب الارقام ولكن المفروض اصلا انى فى نوعين فى ال Categorical attribute واحد منهم بيكون ordinal والثانى بيكون Nominal يعنى واحد الترتيب فيه مهم والثانى لا فى حاله ال Nominal مهم جدا نتعامل معاه عن طريق ال OneHotEncoding وده لان بيدى 1 لل category الى ظهر والباقي ببديلة 0 وهنا مفيش اى يعنى لكل category بيكون فى عدد ال category الى فى الداتا واحد منهم ب 1 والباقي ب 0. من هنا بيظهر حاجة تانيه وهو انى بيكون عندى Sparse Matrix الى هو الماتركس معظمها اصفار فى الحالة ديه لو بتستخدم ال sklearn هاتلاقى بتعلها compress بانها تعمل save فقط لل location بتاعت ال ones وده بيساعد فى حالات ال categories الكثير لمتغير واحد وبيحافظ على الميمورى وفى حالة انك عايز تشوف ال matrix بتحولها عن طريق toarray method.

https://www.linkedin.com/posts/abdelrahman-rezk_machinelearning-ai-d-atasience-activity-6749750455858987008-q_kw

2- Post 2

فى بعض الأحيان لما تيجى تقسم ال dataset ل train and test عن طريق مثلا sklearn method زى train_test_split إنتا مش بتاخذ بالك هل التقسيم ده غطى كل ال Categorize المختلفة فى الداتا ولا فى الحالة ديه إحنا ممكن نشوف اى categorical attribute وتعتبر تقسيمك هايكون عليه هو لو مفيش بيقا هاتشوف من ال continuous attributes وتبتدى تشتغل عليه طيب هاتعمل فيه ايه هنا بقا فى method اسمها Stratified Random Sampling ودية ببساطة بتخليك تحصل على test set and train set بتمثل كل ال population الى عندك.

https://www.linkedin.com/posts/abdelrahman-rezk_dataanalysis-datasets-random-activity-6749339815063429120-a2mr

3- Post 3

التلات صور دول بيترسموا بنفس ال method بس مجرد ال parameters المستخدمه خلت في وضوح اكثر لعلاقات اكثر بين ال two attributes ونقدر من خلال ده نحصل علي بعض ال insights الي مكنتش موجوده في ال plot الاول وأني ازاي اقدر توظف ال parameters ديه مع ال attributes بيفرق كثير في نتيجته ال plot في الاخر سواء في ال size of points or colors كل حاجه لو حاولنا اننا نوظفها صح وخدنا فيها شويه وقت هانلاقي النتيجة في الاخر مختلفه وال insights or ideas الي انتا خدتها أو اكتشفتها هاتكون اكثر وهاتوفر عليك كثير بعدين وطبعا ممكن ده كل يتوظف عن طريق برنامج زي Tablue بدل الكود برضه وهايكون اسهل كثير

https://www.linkedin.com/posts/abdelrahman-rezk_handson-machinelearning-dataanalytics-activity-6772939446686949376-0d--

4- Post 4

ال plots دائما بتديك انطباع سريع وده لان المخ بيكون سريع في التقاط المعلومه من الصور فمثلا حاجه زي ال histogram plots بتساعد في انك تعرف ال ranges وال points الي بتمثل اعلي نسب في كل attribute لو كنت عملت علي الداتا كلها وكمان النسب ديه ممكن تخليك إن بعض ال attributes ممكن بعض القيم فيها أو ranges تعتبرها outliers بسبب بسبب اني نسب تواجدها قليل جدا بالنسبه لقيم تانيه + بيخليك تشوف ال scales المختلفه ويديك ملحوظه انك ممكن تكون محتاج feature scaling ولو كان ال target عباره عن continuous output يخليك تسال لو عايز تعرف هل ينفع القيم بتاع الموديل في التوقع تعدي ال maximum الحالي في الداتا ولا لا

https://www.linkedin.com/posts/abdelrahman-rezk_%D8%A7%D9%84-plots-%D8%AF%D8%A7%D8%A6%D9%85%D8%A7-%D8%A8%D8%AA%D8%AF%D9%8A%D9%83-%D8%A7%D9%86%D8%B7%D8%A8%D8%A7%D8%B9-%D8%B3%D8%B1%D9%8A%D8%B9-%D9%88%D8%AF%D9%87-%D9%84%D8%A7%D9%86-activity-6772594565162549248-1rg3

5- Post 5

بعض ال methods الي بتستخدمها زي ال head بتقديم انطباع أولي عن شكل الداتا بعد كده حاجه زي ال info تبتدئ تخش معاك في تفاصيل اكثر زي كل column ايه هو وال data type وهل بيحتوي علي nan-val ولا وكمان حاجه زي حجم الميموري بتاعت الداتا الي ممكن تخليك تفكر في

جانب ثاني وهو هل هانستخدم معاها حاجه زي ال mini-batch ولا stochastic models عشان يقدر يحصله fit في الميموري بتاعتي وهكذا كمان حاجه زي ال value_count الي بتخليك تشوف ال category attribute فيه كام class وكل class بيحتوي علي كام مثال من ال data بناء عليه ممكن في عمليه ال split data تخليه يكون حاجه base انك تقيم عليه بحيث يكون فيه نسب قريبه من توزيع ال classes المختلفه في ال train and test ولو حاجه زي ال numbers و feature مهم ممكن تحوله ل category تقسم عليه لو كان مثلا ظاهر في حاجه زي ال correlation أنه بياثر جدا علي ال target وطبعا الارقام مع ال describe هاتديك شويه statics حلوه وكمكان ممكن تاخذ بالك اني بعض ال attribute رغم أنها ارقام بس ال ممكن تكون categorical لما تلاقي مثلا النسب قريبه من بعض بي حاجات زي ال first & third quartile and median قريبين جدا من بعض وال std من ال mean عامل ازاي كل ديه حاجات اظن لو شوفناها بشكل مختلف الدنيا هاتختلف

https://www.linkedin.com/posts/abdelrahman-rezk_machinelearning-artificialintelligence-regression-activity-6772238531810123776-t3dq

6- Post 6

من الحاجات المهمه وإنتاج بتفكر في حل تشوف ايه اصلا ال objective بتاعت ال system الي انتا بتشوفله حل لان ده بيخليك تقدر تعمل frame المشكله بشكل صحيح وكمكان ساعتها تعرف هل الموديل بتاعك هيساعد في حل المشكله وهايحقق ال objectives الخاصه بال system ولا لا وأنتك تسأل كمان عن ال system الحالي نتيجه ايه ده هيساعدك انك تقيم شغلك وكمكان ممكن يدريك insights تساعدك وانتا شغال + بعض الافتراضات منك ممكن يخليك تاخذ بالك من بعض الزوايا الي ممكن لما تمشي وراها تلاقي نتائج كويسه زي اني ممكن المشكله اصلا تكون عشان هي توقع اسعار فانتا تفتكر أنها continuous output بس لما تبتدئ تسال ممكن يطلع في الاخر اني ال system في الاخر بيبطلع low price or medium or high وبناء عليه بيتاخذ قرار معين فتلاقي اني اصلا ال problem في frame ثاني غير الي كنت بتفكر فيه وهكذا.

https://www.linkedin.com/posts/abdelrahman-rezk_machinelearning-ai-activity-6771823703425634304-90ql

7- Post 7

إرتباط ال x with y بتاع زمان بيظهر كثير جدا في توضيح العلاقة بين ال attributes الي عندك وال target لو كنت بتتعامل مع continuous output وعندك كثير من numbers attributes ف إنتا محتاج تشوف ايه اكثر attributes فيهم بتأثر على ال target ده وكمكان تشوف العلاقة بينهم هل هي strong positive or strong negative بمعنى ده لما يزيد ال target بتاعك يزيد ولا العكس طب قوة الإرتباط ديه ممكن تخليني مع بعض ال attributes إنني اتاجهها بدام ملهاش اي تأثير على ال target وممكن إنها تسبب مشكله زي ال under fitting لما يكون عندك attributes كثير بس معظمهم بيكونوا اقل تأثير على ال output بتاعك كمان ممكن تحاول إنك تشوف تأثير ال

different attributes على ال target مع بعض ممكن يخليك تربط بينهم بمعادلة وتنتج feature جديد أو لو كان 2 attributes ليهم نفس التأثير فهل هنا محتاجين نتجاهل واحد فيهم ولا لا وهنا ال sklearn بتوفرلك إنك تشوف التأثير على attribute معين بالنسبة لكل ال attribute وكمال بتوفرلك scatter matrix لو إنتا عايز تشوف العلاقة على رسومات بدل من مجرد أرقام

https://www.linkedin.com/posts/abdelrahman-rezk_machinelearning-data-analytics-correlation-activity-6774399910549762048-MTSQ

8- Post 8

من ضمن الحاجات الخاصة بال Features Engineering هو موضوع ال attribute combination لأن بيبقا فيه بعض ال attribute مرتبطة ببعضها ف أنك تطلع علاقة بينهم ده بيزود عندك عدد ال attributes الى ممكن تخلى الموديل بتاعك يطلع نتائج أحسن وده بيرجع لمدى إرتباط برضه ال attribute الجديد ده بال target output بتاعك فبعد أما تحاول تعمل بعض ال combination تقدر تشوف ال Correlation دلوقت بالنسبة لل Target بقت عاملة إزاي وممكن جدا تلاقى إرتباط ال Attribute الى إنتا ضفته ده من علاقة معينة أشد إرتباطاً بالنسبة لل Target من ال Attributes الى عملت منها العلاقة سواء كان ال Correlation ده strong positive or strong negative فى كلتا الحالتين إنتا مستفيد بداتا جديدة فى مقابلها ممكن تستغنى عن Attribute موجود أصلا فى ال Data ولكنه أقل تأثيرا على ال Target

الصورتين دول بيوضحوا قبل وبعض ال Combination الى عملناه من بعض ال Attributes.

https://www.linkedin.com/posts/abdelrahman-rezk_dataanalytics-attributes-combination-activity-6775417290562310144-RTqd

9- Post 9

موضوع إنك تخلى الحاجات automated ده بيوفر عليك جدا قدام وحاجة زى انك تهيه الداتا بتاعتك لل model من خلال كل الخطوات الى عملتها سواء فى ال cleaning أو إنك تعمل Attribute Combination أو فى إنك تهندل ال Categorized attributes كل ديه حاجات وغيرها كتير على حسب الداتا الى بتتعامل معاها وعلى حسب إنك بعض أما عملت شوية analysis على الداتا لقيت إنى بعض ال features ملهاش تأثير على ال output بتاعك فقررت إنك تشيلها أو إنك تحاول تخرج منها علاقة مع متغير تانى كل ديه أفكار وخطوات وفيه كتير منها بيكون بعض الإفتراضات الى إنتا بتحاول تمشى وراها لأنها ممكن تكون سبب لنتيجة أحسن مثلا إنك تفترض بناء على بعض ال analysis إنى ممكن بعض ال outliers فى متغير معين ممكن يكون ليها السبب فى إنى يحصل misleading للموديل بتاعك وغيره كل ده وغيره كتير وإنتا شغال على داتا بيتبندى يظهر وإنى كل حاجة تعملها بشكل عبثى شوية وتقول هاتبقا تظبطها بعدين ده مش صح كنت قرأت قبل كده مقولة لحد إنى 90 % من المشاكل فى الكود بتطلع من جملة هابقا اظبطها بعدين ... فالأحسن والله أعلم

إنك تخلى كل حاجة automated سواء عن طريق function أو class معين وفى الآخر تمشى كل حاجة منفصلة فى Pipeline خاص بيها ويكون فيه PipeLine مجمع لل Pipelines المختلفة

وڪمان لو ڪل حاجه documented عشان لما ترجعلها محتجش وقت إنك تترجم ڪل حاجه

https://www.linkedin.com/posts/abdelrahman-rezk_dataanalytics-machinengineering-ai-activity-6775867975229587456-cGgK