A1

Perfekt. Meine Software läuft auch. Wir nehmen immer mit beiden Systemen auf, dann haben wir ein Backup. Klassisches Informatiker denken, kein Backup, kein Mitleid. Genau. Und dann können wir auch reinsteigen ins Interview. Genau, also du hast ja gerade schon kurz gesagt, Ingenieurshintergrund, bist aber auch bei der technischen Umsetzung von so KI-Lösungen jetzt irgendwie vertraut. Was sind denn deine arbeitsalltäglichen Prozesse, sag ich mal, bei der Entwicklung und bei der Bereitstellung von KI-Produkten?

I13

Ja, okay. Also jetzt wirklich in den letzten zwei Jahren sinds jetzt wirklich erstmal fundamentale Analysen in Databricks. Also da kommen jetzt wirklich kaum irgendwie Machine Learning Modelle zum Einsatz. Das war vielleicht ja davor in den Jahren der Fall, dass ich da Scikit-Learn-Modelle verwendet habe, auch was mit TensorFlow gemacht habe. Und das war dann eher sich die Toolboxen anschauen das heraus… also das erste Problem ist, was dann auch schon in Richtung Labeling vielleicht zielt ist, welches Modell nehme ich denn? Es gibt ja schon so eine große Bandbreite. Zum Teil ist das ja schon schön geclustert, wenn man jetzt bei den Toolboxen Scikit-Learn und so ist. Aber ja, dennoch muss man noch viel lesen und vielleicht auch schon viel vielleicht auch Erfahrung haben. Weil dann, ja, weil man das ein oder andere schon mal gehört hat und vielleicht schon eher weiß, in welche Richtung es geht. Aber das habe ich auch bei Kollegen mitgekriegt, die noch weniger so ins Signalanalysen vertieft waren, dass denen das noch schwerer fiel. Die dann auch auf mich zukamen: Mensch XXXXX welches Modell könnten wir da und da verwenden? Also das ist schwierig. Aber nochmal auf die Frage eingehend, also wie der Alltag so aussieht. Damals, wenn es jetzt auf ML-Modelle zielt, Standard-Toolboxen verwenden, erstmal Scikit-Learn, sich zurechtfinden in diesem ganzen Baum, was überhaupt möglich ist und dann erstmal, ja, austesten, wie gut das funktioniert. TensorF low war noch eine andere Geschichte. Da habe ich jetzt mal ein Beispiel gemacht, wie diese Variational Autoencoder. Da ist es dann, also da wusste ich schon genau, ich werde dieses Modell verwenden. Da war es dann halt, gutes Beispiel finden und umbauen, anpassen.

A1

Ja, genau. Also du hast ja gerade schon so ein bisschen das Labeling angeschnitten. An der Stelle einmal kurz, habe ich gerade vergessen, bei der Einleitung noch kurz als Info für dich. Also wir haben natürlich auch ganz konkrete Forschungsfragen im Hintergrund, die wir mit den Interviews ausloten wollen, aber das ist nicht unbedingt das, was wir jetzt in den Interviews eins zu eins stellen werden. Es ist halt wirklich eher so ein offenes Gespräch darüber. Klar, die Leute, die im Interview mit uns sind, die sind schon ein bisschen gebrieft, die haben zumindest die Webseite gesehen, vielleicht haben die sich auch die Grafik da angeguckt und so. Das heißt, die wissen grob, worum es geht. Aber ja, für uns ist halt wirklich hilfreich. Wir wollen halt wirklich rausfinden, wo die praktische Relevanz ist. Genau, aber es gibt keine richtigen, keine falschen Antworten in dem ganzen Gespräch sozusagen. Das nur nochmal vorweg. Genau, also ein Problem ist auf jeden Fall halt überhaupt dieses Verständnis, das technische Verständnis, was passiert da in Machine Learning? Was sind Modelle? Was tun diese Methoden? Was sind die Vor- und die Nachteile? Das ist so ein bisschen das, was du ja gerade schon erklärt hattest im Bereich Modellsuche oder Methodenauswahl. Das ist am Ende ganz viel testen und ausprobieren. Genau, und wir haben schon kurz darüber gesprochen, aber hast du so konkrete Ideen wie wir als Forscher in dem Bereich das vielleicht einfacher machen können also diesen Technologietransfer und dann wie du gerade sagtest auch Leute abzuholen die vielleicht noch weniger technisches Verständnis von der Materie haben?

I13

Das ist echt erstmal eine gute Frage. Also wenn ich erstmal so denke, wie bin ich denn da dran gekommen? Das war dann wirklich die Vorlesung hören, darüber war dann einst schon viel bekannt und dann ging halt so die Verzweigung nach Vertiefung nochmal, wenn man wirklich bei Scikit-Learn zum Beispiel nochmal die Beschreibung liest, wie man es jetzt noch besser machen kann das ist dann immer schwierig. Also weil da ist es, also eigentlich muss es ja irgendwie Clustering sein. Erstmal, welche Methoden gehören zueinander? Es muss irgendwas Aufgeräumtes sein und in dem jeweiligen Cluster muss dann nochmal separiert werden, warum ist Methode A besser oder schlechter als Methode B, für welchen Anwendungsfall. Also irgendwie müsste man, Grobanwendungsfall überhaupt die Cluster, Regression, Clustering als Beispiel, dann bin ich im Clustering und jetzt unsupervised, supervised, also immer weiter runter. Dann hat man die unsupervised-Methoden, und dann auch wieder separieren, immer wieder verfeinern und die Unterschiede aufzeigen. Vielleicht müsste man, ja das sehe ich auch oft in meinem Alltag, immer von dem Anwendungsfall aus denken, damit es nicht so abstrakt wird, welche Anwendungsfälle gibt es, was sind die Hauptanwendungsfälle, vielleicht kann man sich die Top 5 nehmen oder sowas und gucken, wie man das dann auf die dann aufdröselt. Diesen Baum, den ich jetzt gerade probiert habe zu beschreiben, der sich feiner verästelt. Und dann halt so eine Separation aufzeigen. Bei mir zum Beispiel, ich glaube, jeder hat auch andere Anforderungen. Setzt man jetzt irgendwie so ein Large Language Model ein, das man selbst trainieren will, ich weiß nicht, XXXXX hat man gesagt hier, dass von GPT, also von ChatGPT, kostet wohl 20 Millionen, um das einzutrainieren, das ist nicht meine Anforderung. Ich mache ja hier nur diese kleinen Modelle und da dauert ein Training, weiß nicht, eine Stunde höchstens oder sowas.

A1

Ja ja, genau.

I13

Deshalb brauche ich zum Beispiel nicht auf Kosten gucken. Für mich ist dann wirklich, welche Auswahl der Algorithmen ist denn die richtige? Da ist vielleicht die Auswahl klar. Man will einen Large-Language-Model, aber dann sind die Anforderungen wieder andere. Wie lange trainiert es, wie politisch korrekt sind die Antworten? Also ich glaube, man müsste gucken, was sind die Hauptanwendungsfälle? Welche Kriterien spielen dann pro Anwendungsfall eine Rolle? Hier gerade auch das Beispiel Kosten bei Large-Language-Model. Bei mir wieder vielleicht das richtige Modell zu finden in dieser Vielzahl dieser kleinen Modelle für Clustering oder Regression. Ja, ich glaube, das müsste man herausarbeiten und darauf ja, und dann halt schön darstellen. Eine Methode vielleicht das Labeling, auf der anderen Seite würde vielleicht auch einfach, ich denke wieder an SciKit-Learn, eine aufgeräumte Website, die immer weiter verästelt, ja auch hilfreich sein.

A1

Ja, also ein großes Problem ist auf jeden Fall die Vielseitigkeit, die da sowohl in den Lernaufgaben als auch in der Methodik halt herrscht. Also wir haben da irgendwie jetzt ganz schön als Beispiel, von Sprachmodellen weg geht zu Bildmodellen, da haben wir halt sehr viele Jahre, die wir schon irgendwie versuchen, gute Bildmodelle zu erstellen, angefangen von MNIST und dann waren da Random Forest eine Zeit lang ganz gut, dann kam halt irgendwie Convolutional Neural Networks und dann vor vier, fünf Jahren wurde halt alles nochmal komplett von hinten aufgeräumt mit Vision Transformern, die wieder halt irgendwie Attention Layers eingebaut haben und damit auf einmal alles in Schatten gestellt haben, mit weniger Parametern teilweise bessere Ergebnisse liefern. Aber alleine das zu verstehen, welche Möglichkeiten es da gibt, das ist ja alles extremst nicht trivial. Irgendwie die Situation ist, dass immer mehr Leute das halt in der Praxis nutzen wollen. Und klar, also Sprachmodelle zu trainieren, da reden wir über komplett anderen Aufwand in den Daten und im Training halt irgendwie als bei wir haben hier einen tabellarischen Datensatz von irgendeinem Industrieprodukt oder irgendwelchen Verkäufen so 10.000 rein, paar Features oder schmeißen wir halt irgendwie mal so ein paar Standard Classifier drauf oder Regressoren. Das ist ja ein komplett anderer Anwendungsfall, als wir haben hier irgendwie eine Million Bilder oder eine Milliarde Bilder und wollen da drauf irgendwas machen oder halt Sprache so. Würdest du denn so persönlich die Webseite von SK Learn als geeignet ansehen, um das zu lernen? Also ich meine, es geht damit, klar, offensichtlich. Aber hast du das Gefühl, die ist gut strukturiert und man kommt da gut durch oder hat auch das so eine ziemliche Lernkurve erstmal gehabt für dich?

I13

Ja, teilweise sind die Inhalte, die haben ja manchmal also erstmal würde ich sagen, ist schon ganz gut Schullote würde ich sagen gut, ja gut minus zwischen gut und befriedigend irgendwas ist schon gut, dass die Methoden schon gut geclustert sind, finde ich aber man ist direkt, man ist recht schnell, also du wählst die Regression aus und schwupps, hast du auf einmal alle Methoden, die sind da nicht nochmal weiter subgegliedert, sage ich mal da hast du dann auf einmal 20 Methoden ja gut, doch, ich glaube da gibt es nochmal auch Überschriften, das hilft einem noch ein bisschen, aber man muss dann noch recht viel lesen und was vielleicht auch noch ein bisschen, vielleicht ja, ich weiß gar nicht, ob man es anders darstellen kann aber was vielleicht etwas verwirren kann ist wir haben halt die Methodenbeschreibung, also die Modellbeschreibung auf einer Website und dann nochmal die mathematische Erklärung auf einer anderen das ist nicht so alles in einem das finde ich manchmal ein bisschen verwirrend, wie man da so ein bisschen rumspringen muss, aber an sich finde ich das schon ja ganz gut und noch eine Sache, die mir einfällt, neben dieser ganzen Modellauswahl was bei mir noch wichtig ist, so als dieser Basic Functions Anwender auch jetzt nochmal die Operationalisierung denn das Thema hat, hatten wir ja auch mit dem Basic Smart Sensor, das wäre für mich auch nochmal wichtig, so als Randthema streue ich jetzt einfach nochmal mit rein wie gut lässt sich das operationalisieren was gibt es da schon, also da auch nochmal so Einstiegshilfen

A1

Ja, auf jeden Fall, weil auch das ist ja komplett unterschiedlich, ob man da jetzt von Sprach- oder Bildmodellen redet oder Basisalgorithmen und so und welche Library Dependencies man hat und so. Das ist absolut nicht trivial, wie gut sich das einbetten lässt in bestehende Systeme. Wie ist das, wenn du mit Kollegen über das Thema redest oder mit Leuten zusammenarbeitest, die sich noch deutlich weniger in der Thematik auskennen? Hast du das Gefühl, dass man da manchmal so Kommunikationsprobleme hat? Ist das generell eine Technologieakzeptanz oder eher eine Ablehnung da? Kannst du in dem Bereich irgendwie noch was sagen?

I13

Also ich finde doch, dass viele Leute auf jeden Fall die Power, die dahinter steckt, die sie schon sehen, jetzt nicht mit der Revolution, hier mit ChatGpt, das ist ja wirklich der Hammer. Also ich glaube, da wurden auch viele nochmal zusätzlich überzeugt. Aber ich glaube, man kann es ja recht schnell anwenden. Ein paar Zeilen und du hast ja schon ein Regressionsmodell fertig. Also ich glaube, dass viele Entwicklungsingenieure das vielleicht schnell anwenden können. Ich finde aber, wenn man dann in die Tiefe geht, dass da dann doch noch ein bisschen das Verständnis für gewisse Sachen fehlt. Und ich glaube, dass viele sich dann glaube ich auch schon sehr sicher vielleicht wägen. Man guckt sich ein paar YouTube-Videos an und denkt, dass man dann schon sehr viel kann. Aber ich glaube, das geht ja noch sehr viel tiefer, die Thematiken. Und ja, also das Verständnis drum, was da noch hinter steckt, so würde ich vielleicht sagen. Auf der anderen Seite, das kommt mir jetzt auch in den Kopf. Vieles ist aber auch in dieser Machine Learning Welt ja auch wirklich anwenden. Und vielleicht muss man mit GRIT Search drüber laufen und man hat vielleicht auch schon gute Ergebnisse. Also ich führe gerade so quasi so ein Brainstorming mit mir selbst durch. Vielleicht, ich frage mich gerade, wie viel man vielleicht auch wissen muss. Zum einen Teil ist es natürlich immer gut, wenn man jetzt Entwickler ist, zu verstehen, wie das Modell funktioniert. Aber ja, da müsste man sich auch nochmal die Frage stellen, inwieweit muss ich es denn wissen? Und inwieweit, ja, nur für die Anwendung. Ich muss es ja nicht weiterentwickeln. Das ist vielleicht auch nochmal so ein interessantes Feature, wenn man irgendwie so ein Inhaltsverzeichnis für ML-Modelle erzeugen will. Wie weit muss ich gehen, wenn das Thema nur Anwendung ist? Also ein bisschen muss man natürlich erklären. Also das, ja, aber das sehe ich oft. Also dass die Leute schon so ein Grundverständnis haben, kriegt man ja sehr schnell vermittelt bei YouTube. Mit ein paar Zahlen kriegt man schon erste Modelle hin. Ich glaube, das ist schon ganz gut. Dann, das weiß ich auf jeden Fall aus Diskussion mit Kollegen zu einem Projekt. Ein Kollege sagte auch, Mensch, die Vielzahl der Modelle, ich weiß gar nicht, was ich nehmen soll. Ich nehme mal einen Tipp. Aber das hatten wir schon eingangs. Also die Orientierung gar nicht zu verlieren. Aber so sehen die Leute wirklich die Power. Ich merke das auch im Unternehmen, dass das jetzt auch schon auf den höheren Chefetagen so angekommen ist, so lotet mal aus, was könnte noch gehen und so. Ich weiß, XXXXX, der entwickelt weniger. Der will halt eher anwenden. Guckt schon, was ist auf dem Markt. Was kann man bei XXXX schnell implementieren? Und so von den Chefetagen, aber auch, hey, Ingenieure, guckt mal, was können wir denn, was gibt es hier für Potenzial? Was könnten wir noch entwickeln? Also da finde ich schon, ist ja auf jeden Fall eine Affinität. Man sieht Potenzial. Ja, da will man weiterhin.

A1

Ja, okay. Ja, das ist ganz interessant, weil wenn man so die Bevölkerung fragt im Schnitt, jetzt nicht unbedingt nur Leute, die in der Firma was entwickeln, dann ist da natürlich immer noch eine sehr starke Skepsis auch bei der Technologie. Ich glaube, da ist aber dann auch wieder das Daily Business, was XXXX betreibt, da ist den Leuten wahrscheinlich relativ egal, warum die Wasserpumpe das tut, was sie tut. Wogegen, wenn es jetzt um, weiß ich nicht, Krebsrührerkennung geht oder so, die Leute schon deutlich mehr Muffensausen kriegen, wenn sie lesen, dass da KI eingesetzt wird.

I13

Ja, okay.

A1

Weil da eine andere Form von Kritikalität dabei ist oder auch im Bereich autonomes Fahren ist ja diese moralische Frage danach, sollte eine Maschine irgendwelche Fahrentscheidungen treffen dürfen. Da ist auch sehr viel Unmut in der Bevölkerung an ein paar Stellen. Aber genau, also wir nehmen das erstmal so hin. Ist auf jeden Fall cool zu hören. Also ich bin ja offensichtlich auch irgendwie beweis dafür, dass XXXX Bock auf KI-Transfer und irgendwie die KI-Nutzung im Unternehmen hat, sonst hätten wir ja die Partnerschaften nicht.

I13

Ja, du hast recht, du hast mich nochmal angestubst. Ich habe jetzt sehr aus meiner Blase heraus gedacht. Du hast auf jeden Fall recht. Auch Freunde, die wirklich weiter weg sind von der Technik, klar, denen bereitet das vielleicht auch nochmal Sorgen, dass man eben Cyborgs baut oder jetzt politisch kriegt man das ja auch immer mit, so viele Fake News passieren. Ich habe gerade eben noch hier beim Essen einen Bericht gelesen, dass Aussiedlerbote.de Nachrichtenmagazin angeblich mit 23 Redakteuren und alles nur KI-Fake. Tagesschau hat die dann, also Recherchenetzwerk, ich weiß nicht, NDR, MDR und was weiß ich, hat dann den Typen damit konfrontiert. Ja, der hat dann doch nochmal sein Impressum ein bisschen geändert und gesagt, nee, wir machen hier KI-Forschung für Reportagen und sowas, aber nur, glaube ich, weil er aufgeflogen ist. Also ich glaube, das macht den Leuten Angst. Also da hast du schon recht. Die Beeinflussung, die dadurch passieren kann, ja.

A1

Ja, also ich meine auch mit Recht, man muss das auch kritisch sehen, man muss jegliche technologische Entwicklung auch immer kritisch hinterfragen, aber ich meine ja, wir machen ja auch die Umfrage eben mit Anwendern in der Industrie und halt Firmen und Leuten in Firmen, die halt KI-Produkte entwickeln und anbieten wollen, insofern ist es auch richtig, dass du da aus der Sicht halt irgendwie das schilderst und da ist eine Technologie-Affinität und ein Interesse und eine richtige Begeisterung dafür da, ist doch gut, also als Forscher auch. Also logisch. Ja, wir leiten mal über in den eigentlichen Hauptteil. Labeling, du weißt schon grob, worum es geht, auch wenn wir uns, glaube ich, noch nie drüber unterhalten haben, aber du hast die Webseite halt gesehen und XXX XXXX XXXX XXXXXX XXXXXXXXXXXXX XXXXXX und so. Also ich teile dafür erstmal meinen Bildschirm, das ist, glaube ich, das, was ich damit anfange. So, du solltest, ihr solltet meinen Bildschirm sehen. Bevor ich jetzt hier im Detail erkläre worum es geht oder was der Sinn des ganzen ist. Kannst du mir einfach mal so eine spontane Reaktion geben auf das was du siehst? Was sind deine Gedanken dazu? Einfach frei raus.

I13

Erstmal muss ich mich zurechtfinden. Also ABC, das sticht direkt ins Auge, das ist gut. Ich sehe eine Assoziation zu Sachen, die ich schon kenne. Labeling von technischen Produkten, Waschmaschine etc. Dann oben, wo ein AI-Label ist, dann klar, darunter die Sachen. Die wirken erstmal für mich ein bisschen verwirrend. MobileNetV3, Small. Wahrscheinlich geht es da schon um ein Beispiel für ein Produkt. Aber das würde mich erstmal verwirren. Und unten würde ich… ja das würde ich dann auch, diese vier Punkte, erstmal deuten als Benchmark-Features, vier Stück. Batterie, wahrscheinlich Powerdraw per Inference. Also pro Modellanwendung, wie viel Strom gezogen wird. Dann Genauigkeit des Modells. Corrupted Robustness. Achso okay vielleicht wie robust das ist. Also für Fehlqual-… je nachdem, was das für ein Datensatz ist. Wie zuverlässig ist es? Ja, finde ich ein bisschen schwierig, dann in Abgrenzung zu Accuracy. Und Running Time per Inference. Okay, das ist auch klar. Wie schnell berechnet ist das? Okay, das ist erstmal klar. Ja. Also nur oben, das verwirrt mich. Aber wahrscheinlich, wenn ich wüsste, welches Produkt das ist, wenn das irgendwie eine XXXXXXXXX wäre, dann hätte ich mich da auch zurechtgefunden.

A1

Ja, genau. Als Hintergrund zu dem oben. Also es geht tatsächlich um ein konkretes Bilderkennungsmodell und zwar Name des Modells ist MobileNet V3 Small. Das wurde trainiert für ImageNet oder auf ImageNet. ImageNet ist einer der bekanntesten großen Bilderkennungsdatensätze. 1000 Klassen pro Klasse, 1000 Bilder mit einer im Verhältnis recht hohen Auflösung, weil wir irgendwie 300 mal 300 Pixel oder so haben in dem Datensatz. Und MobileNet V3 ist insbesondere ein Modell, was auf Mobile vor uns halt irgendwie laufen kann. Also es wurde für so einen gewissen Edge-Einsetzungszweck trainiert. Relativ aufwendig im Training. Das braucht tatsächlich ein paar Wochen auf einer einzigen GPU, bis das austrainiert ist. Kann dafür dann aber mit relativ wenig Parametern sehr effizient halt auch auf kleinen Geräten laufen. Deswegen heißt es MobileNet. Wenn man die QR-Codes scannt oben, dann kommt man auch zu dem Paper von dem Modell und kann sich da noch mehr Informationen holen. Das so ein bisschen als Erklärung, dass du es oben einordnen kannst. Hast du das Gefühl, dass so eine Darstellungsform in irgendeiner Form irgendwie eine Brücke zu deiner Arbeit hat?

I13

Also ich glaube, zu meiner weniger. Ich könnte mir vorstellen, aus XXXXXX Sicht, der Modelle anwendet, der die nicht entwickelt, da schon, da bin ich halt weiter weg.

A1

Also du würdest praktisch eher Modelle schaffen, die man dann labeln könnte?

I13

Ja, bei mir ist es ja wirklich eher so Regression, Klassifikation, Variational Autoencoder anwenden, das ist, ich finde, ja, das sind immer so kleinere Modelle, das ist nicht so was Großes, schon Fertiges, was man jetzt einkauft und irgendwie gezielt anwendet, ne?

A1

Ja, das ist völlig richtig, aber ich habe tatsächlich auch mal ein Paper geschrieben, wo ich das für so Standard-Classifier gemacht habe, weil im Endeffekt, ob du jetzt ein neuronales Netz labelst, also wo machst du den Trade-off? So ein Random Forest mit 1000 Trees hat auch so viele Parameter, dass du in der Größenordnung von einem neuronalen Netz kommen würdest. Klar, bei einer linearen Regression mit 10 Features oder so, das ist wirklich, das als Modell zu bezeichnen sozusagen oder auch als KI zu bezeichnen, ist vielleicht ein bisschen stretched, aber es ist halt sehr schwer, da den Cut-off zu machen und sagen, das ist eine klassische statistische Methode und das ist jetzt eine KI. Das ist halt total fließend. Insofern, du könntest ja genau die Messwerte unten, könntest du genauso auch für Standard-Classifier oder Regressoren angeben. Die haben dann vielleicht ein bisschen andere Metriken, aber an sich kann man das trotzdem machen.

I13

Ja, ich denke, das sind jetzt genau die Sachen, aus dem Anwendungsfall heraus gedacht. XXXXX, eher Anwender von größeren Modellen, ich glaube, da passt das. Wie lange läuft das, wie robust ist es? Viele Nutzer nutzen das. Robustheit, Daten sind unterschiedlich. Es läuft die ganze Zeit durch, wie viel Leistung zieht es. Bei mir eher ist das, was wir vorher gesagt haben, boah, wie finde ich mich zurecht in Scikit Learn? Ich bräuchte wieder so als Entwickler einfacher, ich würde das auch nicht AI bezeichnen, bei mir ist das Machine Learning. Einfache Machine Learning Modelle, da würde ich mir andere Features wünschen. Vielleicht wirklich Klassifikation, dann Klassifikation von numerischen oder von Klassen. Es gibt ja auch Inputs, die halt nicht einfach Zahlen sind, sondern wirklich, wie soll ich sagen, kategorische Variablen. Dafür geeignet, also ich, bei mir ist eher sowas der Fall. Dann vielleicht noch Embedded geeignet, nicht Embedded geeignet, vielleicht wie gut operationalisierbar in Databricks. Ich bräuchte, glaube ich, andere Merkmale. Also das ist das, was ich eingangs meinte, glaube ich. Je nachdem, was der Anwender macht.

A1

Ja, andere Merkmale ist aber ein total guter Punkt, weil das ist natürlich auch, also wir haben die ja jetzt erstmal entworfen, so diese Label und das ist natürlich, also wir machen ja jetzt keine, im Detail keine User-Study dazu, aber es gibt schon verschiedene Arbeiten, die auch sagen, dass jegliche Form von Erklärbarkeit für KI und Co., das muss immer irgendwie so ein bisschen dialogartig sein oder zumindest die Möglichkeit haben, angepasst zu werden, weil halt unterschiedliche Zielgruppen auch unterschiedliche Darstellungen brauchen. Das ist ein bisschen anders bei der Analogie, die du vorhin gezogen hast zu den Energielabeln, die wir in der EU haben, weil das richtet sich nur an Endanwender und da muss nur eine Information drauf und die treffen dann ihre Entscheidungen gegeben, Budget, was ist das energieeffizienteste Modell und fertig. Es gibt nur, da wird versucht, das auf ein Gütekriterium runterzubrechen, was natürlich irgendwie auch ein bisschen schwierig ist. Macht ja einen Unterschied, ob man so einen Kühlschrank in einer Dachgeschosswohnung oder im Keller betreibt eigentlich, aber es ist diese Komplexität der Entscheidungsmöglichkeiten, die wird versucht, da völlig wegzuabstrahieren. Als Entwickler oder als Programmierer auch in dem Moment, du bist an viel mehr Detailzeit interessiert. Du willst ja keine abstrakte Darstellung, Güte A, B, C, D, E, sondern du willst ja wissen, wie du gerade gesagt hast, wo wird das eingebettet, wie operationalisierbar ist das, wie ist die, keine Ahnung, wann war der letzte Commit? Also wenn eine Library oder so ist dann viel interessanter, wird die noch aktiv weiterentwickelt oder ist das Ding schon deprecated oder in einem Jahr deprecated? Aber als Entwickler bestellst du ja oder du stellst ja am Ende auch deine Lösungen vor, die du entwickelt hast. Also klar, im Prozess des Entwicklens brauchst du tiefergehende Informationen, aber ist das vielleicht was, was im Reporting zum Beispiel interessant wäre, wenn du das halt Leuten erklären willst, die halt eben nicht an dieser Tatsache, dass sie da tiefer interessiert sind oder da überhaupt mitreden können?

I13

Ja, also bis dato habe ich dann immer, vielleicht von Scikit Learn, irgendwelche Bilder genommen oder selbst, also bei den Variational Autoencodern noch ein bisschen vorgestimmt, selbst gezeichnete Bilder, wie so ein Clustering passiert oder, ja genau da oder spezieller noch, ich habe einmal auch so einen Verge-Algorithmus genommen, das ist halt auch noch sowas zum Klassifizieren und das Gute war halt noch, der war auch embedded-fähig, weil wir erstmal gedacht haben, dass wir so einen Algorithmus auf ein Embedded System laufen lassen und da habe ich dann auch nochmal selber einfach nochmal dargestellt, wie so ein Clustering funktioniert, Datenpunkte aufgemalt in einer zweidimensionalen Space, die einen rot, die anderen grün und dann gibt es ein Cluster und dann kommt ein neuer Punkt daneben, das Cluster, lernt dazu weitet sich aus, ich habe es dir ja auf so einer Ebene dann erklärt, Bilder von Scikit Learn genommen, ich weiß nicht, ob ich das mit einem Labeling erklären könnte,

A1

Ja, da geht es ja darum, dass du dann versuchst zu erklären, wie das Modell intern funktioniert. Das geht dann ja in Richtung, nicht ich reporte die Ergebnisse, sondern ich versuche die Funktionsweise der Methode oder des Modells halt irgendwie nachvollziehbar zu machen. Vielleicht anhand von abstrakten Beispielen, so Clustering ist ja immer das Go-To-Ding, das kann man halt super visualisieren, hier zwei Punktwolken und dann, wie bei einer SVM, schiebt sich dann die Gerade immer weiter, sodass das schön linear beides voneinander separiert und dann sagt man, naja, aber was ist, wenn wir jetzt kein linear separierbares Problem haben, dann braucht man höhere Komplexitätsklassen. Das ist natürlich nicht das, was man mit so einem Label machen wollen würde. Ein Label versucht ja nicht zu erklären, wie ein Kühlschrank funktioniert. Also in dieser Analogie, wenn man da jetzt mal bleibt, versucht ja ein Energielabel oder Nutri-Score nicht zu erklären warum es diese Bewertung gegeben hat und wie das Modell überhaupt funktioniert. Die Frage ist halt gibt es da vielleicht einen Kommunikationsbedarf, dass man auf dieser Eigenschaften-Ebene bleibt?

I13

Ja, aber eigentlich hast du recht. Also gut, genau, du hast recht. Das sind zwei verschiedene Paar Schuhe. Das eine ist genau, wie funktioniert das Modell? Das andere ist vielleicht wirklich nochmal, doch, das ist eigentlich auch nochmal gut, so ein Labeling zu haben, wirklich nochmal aus Anwendersicht noch weitere Infos aufzuzeigen. Ich denke nur gerade, also zum Beispiel bei meinem Chef, der ist jetzt weiter weg von Machine Learning, ist ein intelligenter Mann, Physiker, aber schon lange als Manager aktiv. Deshalb, der steckt nicht so tief in den Sachen und ist halt immer skeptisch. Er weiß halt auch, XXXXX, du bist eigentlich Ingenieur, wie gut ist dein Wissen? Deshalb gab es auch damals nochmal XXXXXX XXXXXXXXX XX XXX XXX Xabei wart, um meine Sachen zu evaluieren, damit er sich immer wieder absichert, wie gut ist es denn, weil er es nicht mehr beurteilen kann. Und da würde ein Labeling helfen, weil da wüsste er, das ist abgesegnet von den Profis und die und die Features, die da draufstehen, die sind validiert, da ist was hinter. Nur, es müssten dann die Richtigen sein, die ihn natürlich interessieren. Und ich überlege gerade, was interessiert ihn. Jetzt zum Beispiel bei diesen Basic Smart Sensor Analysen, er sagt halt, alle Analysen sollen halt in zwei Minuten durchlaufen für 50, 100, 200 Basic Smart Sensoren. Deshalb würde ich sagen, Geschwindigkeit interessiert ihn. CPU-Power, ich glaube, das wird ja, vielleicht auch, weil das ist dann wieder Geld. Vielleicht müsste ich dann eine dickere Maschine in den Databricks nehmen, wenn ich dann so, argumentiere. Ja, da auch, Robustheit. Also, ja, also, das ist nochmal zusätzliche Infos vielleicht, die das Modell dann nochmal unterstreichen, was ich ausgewählt habe. Das würde mir dann helfen, ja. Weil es dann von den Profis abgesegnet ist und dann Vertrauen geschaffen wird.

A1

Ja, oder auch einfach so die Vor- und die Nachteile so ein bisschen zu vergleichen. Deswegen, genau, also statt schwer verständlichen Papieren oder abstrakten Implementierungen kann man vielleicht die praktischen Eigenschaften auch so verständlicher irgendwie repräsentieren. Das ermöglicht natürlich auch den Vergleich von verschiedenen Lösungen sozusagen. Und dann könnte man das halt irgendwie der Managementabteilung sagen und sagen, so, das sind die Sachen, die wir ausprobiert haben. Jetzt, liegt die Entscheidung bei Ihnen, was ist wichtiger so? Da kannst du gerne auch nochmal kommentieren, wie gut du jetzt diesen Vergleich hier irgendwie ziehen kannst für dich persönlich.

I13

Also hier. Ja, was sehe ich denn? Genau jetzt sehe ich auch sogar, dass unten ich dachte, das wären einfach nur Farben gibt, Farben der Features, aber jetzt sehe ich, die haben ja auch eine Bedeutung, genau, wie gut das halt ist und jetzt sehe ich schon rechts war jetzt sogar zweimal richtig grün, gelb, einmal rot, das führte dann glaube ich zu einem guten Score A und links sehe ich ja, ist das unten grün schon ein bisschen schlechter, genau, die Zeit ist auch länger, doppelt so lange quasi braucht die Inferenz, die Accuracy ist zwar höher, aber dafür ist auch die Leistungsaufnahme höher, Robustheit auch, aber gut, da sehe ich halt wieder zusammenverrechnet wahrscheinlich auch als Regression die Werte, die reinkamen, die führen dann halt zu einem schlechteren Score, der dann halt, ja, C ist, weil halt die Feature ja, schlechtere Werte abliefern.

A1

Also am Endeffekt ist das eine gewichtete Summe, aber das Interessante ist dabei, die Gewichtung, die ist jetzt hier so gewählt, dass sich die Güte, die Qualitätsgüte sozusagen 50-50 verteilen sollte mit den ressourcenbezogenen Aspekten, also Energieverbrauch und Laufzeit, da geht es ja darum, wie viel Ressourcen verbraucht ein Modell, Robustheit und Accuracy sind halt Metriken, die benutzt werden, um die qualitative oder die Vorhersagegenauigkeit irgendwie einzuschätzen. Man kann natürlich da an der Stelle auch sagen, alles ist egal, solange die Laufzeit unter dem und dem Threshold ist oder uns ist Energieverbrauchs ist uns egal, wir brauchen einfach nur möglichst hohe Accuracy und kann halt diese Gewichtung dann, da geht es dann darum, also da könnte man dann die obere Skala dieses Gesamtbewertungen, die könnte dann im Endeffekt auch eine Managementabteilung gegeben, ihre eigenen Prioritäten oder Constraints irgendwie festlegen und die Modelle halt ranken basierend auf den Metriken unten. Tatsächlich für das Gesamtranking stehen ja auch noch mehr Metriken im Hintergrund, die wir untersucht haben, also statt Top-1-Accuracy zum Beispiel auch Top-5-Accuracy, also war die korrekte Klasse unter den Top-5-Klassen, bei ImageNet bei 1000 Klassen ist das natürlich super relevant, wenn man jetzt ein binäres Problem hat, dann ist das natürlich eine Metrik, die völlig irrelevant ist. Wir haben jetzt hier unten halt nur vier dargestellt, damit das einigermaßen überschaubar bleibt, das könnte man aber natürlich auch dann wieder alles noch anpassen und so. Genau, und dann werden im Endeffekt die absoluten Messwerte, die jetzt unten auch dargestellt sind, werden auf relative Vergleichswerte umgemappt, aus denen ergibt sich sowohl das Farbrating und diese relativen Vergleichswerte werden auch für das Gesamtrating jetzt gerade hier zusammenverarbeitet sozusagen oder zusammen...

I13

Also Relative Vergleichswerte, das verstehe ich jetzt nicht ganz. Also was ich gerade verstanden habe, was ich auch interessant fand, ist, wenn das Management selbst die Gewichtung festlegt und bei Management XXXXXXXXXXXXX kommt hier kommt A und D raus und bei XXXX kommt C und A raus, weil jeder das halt irgendwie anders setzt. Das finde ich eigentlich ganz cool. Auf die Anwendung zugeschnitten. Das habe ich verstanden. Und was ist das, was du mit lokalen… Also absolute Sachen verstehe ich. Das sind die absoluten Features.

A1

Genau, das sind die absoluten Messwerte unten. Diese relativen Werte ergeben sich daraus, dass man praktisch, das wird übrigens im Hintergrund bei den Energielabeln auch so gemacht, daher kommt die Inspiration. Im MobileNet V3 halt jetzt hier zum Beispiel im Vergleich von deutlich mehr Modellen, also wir haben insgesamt 15 verschiedene Modelle evaluiert und MobileNet V3 war das mit der höchsten Geschwindigkeit und dem niedrigsten, also mit der niedrigsten, nein, höchsten Geschwindigkeit, niedrigster Laufzeit und dem niedrigsten Stromverbrauch. Deswegen kriegen die jetzt ein A-Rating und kriegen einen relativen Wert von 1 zugeordnet. Und alle Sachen, die langsamer sind, zum Beispiel links, das hast du ja gerade richtig gesagt, das war ungefähr halb so schnell, ungefähr, würde dann entsprechend diesen relativen Index Wert von 0,5 bekommen.

SPEAKER\_02

Ach so, ja.

A1

Man kann natürlich beliebig schlecht werden, man kann halt irgendwie auch Faktor 100 langsamer sein, dann ist man halt irgendwie bei 0,01 so und nicht bei 0,5. Und so lässt sich das halt dann machen und dann wird auf allen relativen vergleichswerten praktisch diese Bins bestimmt für die Farbcodierung.

I13

Ja, verstehe. Also relativ immer zum besten Modell oder wozu auch immer, zu bestehenden Ergebnissen.

A1

Das ist aber natürlich, in der Tiefe wird das natürlich auch nicht gezeigt oder erklärt, weil das halt wieder, was ist das, das sind wieder Detailinformationen, die Endnutzer am Ende nicht interessieren, aber diese Vergleichbarkeit, das ist halt einer der Hauptpunkte, warum ich das Gefühl habe, dass solche Labels vielleicht praktische Relevanz haben können. Hast du das Gefühl, dass jetzt bei den Metriken unten, so aus deiner Arbeit in dem Bereich, fehlen da für dich Messwerte? Also gäbe es irgendwelche Sachen, wo du sagen würdest, boah, das wäre wirklich, das ist halt wirklich wichtig, dass ich das auf einen Blick sehen könnte? Also wir haben ja gerade schon ein bisschen darüber gesprochen, so ein bisschen diese Charakterisierung, die Einordnung, was ist das für ein Modell, was macht es, in welche Technologie setzt es im Hintergrund ein? Das wäre noch was, was man irgendwie... machen könnte, abgesehen davon noch irgendwelche Sachen?

I13

Ja, also genau. Erstmal, wenn man jetzt so darüber spricht, wo ich jetzt so nicht sehr vertraut damit bin, finde ich, eigentlich sind sie ja wirklich sehr naheliegend. Also wirklich, Rechenzeit ist wichtig. Wie schnell ist es? Dann halt auch wie viel Power, weil Power ist Geld, wenn ich es operationalisiere. Brauche ich eine dickere Maschine bei Databricks, muss ich mehr Geld ausgeben? Robustheit auf unterschiedliche Datensätze ist nochmal gut und dann generell, ja wie gut ist es denn? Das ist schonmal gut. Ich denke nochmal jetzt, aus meiner… Also die Probleme die ich… Wo habe ich pain? Das ist immer bei XXXX jetzt so. Aus dem Schmerz heraus gedacht. Welches Modell soll ich anwenden? Wie finde ich mich zurecht? Das wäre nochmal irgendwie interessant wenn man das irgendwie darstellen könne. Dann aus dem Schmerz heraus gedacht wie operationalisiere ich das? Ich bin jetzt immer in Databricks unterwegs, ich kenne nicht die weitere Welt, Azure Cloud und so, Amazon Cloud und AWS und all diese weiteren Sachen, kenne ich mich nicht aus, ob es dann halt wieder anders ist. Also muss man sagen, ist es gut operationalisierbar in Databricks und vielleicht ist es schlecht in AWS oder umgekehrt, also ob man sowas aufzeigt, vielleicht allgemein operationalisierbarkeit. Gibt es da schon was, das ich gut einbetten kann? Gibt es da irgendwie so ein rapper/wrapper(?), wo ich das Modell reinsetze und mit wenig Handgriffen läuft es? Das ist ein Schmerz. Warte mal, Schmerz ist noch bei mir, das hast du Laufzeit, immer wieder die runterzudrücken. Schmerz sind Kosten, hast du auch im Sinne von Rechenpower. Was ist noch Schmerz bei mir? Ja, erstmal fällt mir da nicht mehr zu ein.

A1

Ja, also ist ja schön, wenn wir da soweit jetzt irgendwie alles erstmal abgedeckt haben. Wie gesagt, was man immer im Hinterkopf behalten muss, ist, wenn man verschiedene Lernaufgaben sich anschaut, kommen die natürlich alle mit unterschiedlichen Metriken auch und es gibt natürlich irgendwie, man könnte sich auch Precision Recall anschauen statt Top-1 Accuracy und so. Also man kann das ja im Bereich Qualitätsgüte, ist es immer super abhängig davon, welche Lernaufgaben man überhaupt hat. Macht man Regression oder Klassifikation oder Clustering oder, ne? Was ich auch ganz spannend finde, mir ist das tatsächlich jetzt irgendwie gerade erst nochmal aufgefallen, also ich weiß, dass das so ist, aber mir fällt es gerade bei den Labels nochmal auf. Viele, also es liegt ja nahe zu glauben, wenn das Modell schneller ist, dann verbraucht es auch weniger Energie und das ist auch meistens so. Aber wir sehen jetzt hier auf den beiden Labels auch sehr schön, dass die Laufzeit sich zwar fast verdoppelt, aber der Energieverbrauch, der verdreifacht sich fast. Ja, stimmt. Es liegt im Endeffekt daran, dass das Modell links auch deutlich größer ist und die GPU mehr auslastet. Das heißt, dass die GPU wird links bei dem Modell einfach mehr Watt ziehen, weil halt mehr, ja, mehr Prozessoren darauf benutzt werden, wenn das Modell irgendwie deployed wird. Wir haben ja in den anderen Interviews nicht im Detail darüber gesprochen, mir ist das nur gerade aufgefallen irgendwie, dass das hier so ein schönes Beispiel ist für Laufzeit ist eben nicht gleich Energie. Ja, korreliert, aber nicht gleich.

I13

Ja, also die einzige Sache ist, man kann sich auch überlegen mit Geld. Wie teuer ist es, das zu betreiben, aber ich finde, das ist ja das ist dann ja gleichzusetzen mit Leistung.

A1

Ja, da bist du auch in diesem Operationalisierbarkeitsproblem. Am Ende ist das halt Python-Code und wie operationalisierbar das ist, hängt absolut davon ab, welche Infrastruktur das Unternehmen nutzt. Ich meine, man könnte das halt auch irgendwie auf einen Raspberry Pi draufpacken, dann kriegt das lokal Daten von einem Sensor, wertet das lokal aus und schickt dann den Rest über irgendeine Post-API irgendwo hin. Das ist ja dann auch operationalisiert. Also die Frage ist halt auch immer, ich finde es sehr schwer mir vorzustellen, wie operationalisierbar ist denn so ein TensorFlow-Modell wie in dem Fall hier oder ein SK-Learn-Classifier. Der ist erstmal nicht gut oder schlecht operationalisierbar. Die Problematik der Einbettung und Anbindung an Daten und Co. liegt dann eher bei dem Unternehmen. Ich glaube nicht, dass das ist, was man über die Methode aussagen kann.

I13

Vielleicht ist das so, XXXXXXX, wie du das vorhin gesagt hast. Ich fand die Idee eigentlich ganz gut, wenn man Eingabeparameter hat. Einer wählt Raspberry Pi aus, der andere wählt Databricks aus, der andere AWS Cloud irgendwie und dann kommt auf einmal ein anderes Labeling raus, ne?

A1

Ja, auf jeden Fall.

I13

[Unverständlich] dass auch jeder was anderes angezeigt kriegt in diesen vier Feldern. Dass es eine Website gibt

A1

Ja, also das Environment hier ist jetzt ja auch irgendwie ein einzelner Cluster-Node mit 8 GPUs. Da ist ansonsten gar keine Abstraktion dazwischen. Da ist eine Python-Instance drauf, da wird das Modell reingeladen und fertig. Das ist natürlich, also ja, auf jeden Fall, dieses, ich nenne das immer Environment, wenn ich darüber schreibe, das muss man irgendwie im Kopf behalten. Und natürlich werden sich die Messwerte unten auch fundamental ändern, wenn man das Environment ändert. Natürlich braucht irgendwie, also so Response-Time von ein bis zwei Sekunden für einen kompletten Batch an Bilddaten. Das ist vielleicht auf so einer fetten GPU-Maschine realistisch, aber auf einem Notebook oder so sicherlich nicht. Ja, genau. Also klar, diese Abhängigkeit von dem Environment hat man. Und da ist dann wieder die Frage, es wäre natürlich cool, eine Webseite zu haben, wo das auf allen möglichen Sachen mal durchgemessen wurde und man dafür die Ergebnisse abfragen kann. Die Alternative wäre, die Leute führen ihre Experimente auf ihrer Infrastruktur durch, kriegen dann hinterher die Labels ausgespuckt von den Sachen, die sie getestet haben. Und das ist so ein bisschen der Weg, wie ich das im Moment fahre. Also die Library, um diese Labels zu erstellen, die ist halt open source. Die findet man auf GitHub und da kann man einfach ein Pandas DataFrame reinladen mit dieses Modell, hat auf diesem Datensatz diese Messwerte und dann kann man daraus dynamisch solche Labels generieren.

I13

Ja, eigentlich wäre das auch cool. Es gibt doch hier diese Website, dieses CPU Benchmark oder wie das heißt, dass jeder User, das man macht und das wird dann zentral in dieser Datenbank gesammelt und dann hat man so ein Ranking daraus, da braucht nicht ihr oder du irgendwie alle Tests zu machen, sondern die Welt arbeitet für einen, aber irgendwie muss es einen Benefit geben.

A1

Ein großes Problem ist, ich habe mir beispielsweise Papers with Code und OpenML mal angeschaut. Das sind ja so Online-Repositories, um einfach so Machine-Learning-Experimente praktisch Open Source zu teilen, damit Leute nachgucken können, wie gut war denn das. Da wird halt Ressourcenverbrauch absolut underreported. Also du hast halt irgendwie 80 bis 90 Prozent der reporteten Ergebnisse sind halt qualitätsbezogen und so Laufzeit und Energieverbrauch wird halt kaum reported. Aber ja, an sich wäre das cool. Das ist auch so ein bisschen die Vision, die ich habe. Kein Bachelor-Studi braucht mehr lokal irgendwie seine Classifier auf seinen Standarddatensätzen durchlaufen, weil das hat jeder so. Ich will nicht wissen, wie oft MNIST schon durchgemined wurde. Oder ADDLT/Adult(?) als Tabulardatensatz oder so. Eigentlich ist das Quatsch, dass das immer noch Leute machen, weil es ist durch das Spiel. Aber ja, trotzdem, es gibt halt kein gutes Repository, um sowas online zu sammeln und irgendwie daraus zu lernen. Genau. Wir haben gerade schon viel über SKLearn gesprochen und Software-Dokumentation und so. Kennst du noch andere Systeme, um über das Verhalten von Machine-Learning-Modellen oder Methoden, die Wörter an der Stelle mal austauschbar genutzt, sich zu informieren?

I13

Ja, also ich habe mir dann auch ein bisschen angeeignet, was ich so toll finde, dass die Community ja wirklich so offen ist weltweit. Da ist ja wirklich, alle Paper, viele Paper auf X… Arxiv oder wie das heißt, stehen da zur Verfügung. Ich meine, da habe ich nicht viel rumgelesen, aber ich weiß, dass die selbst mit Quellcode da zur Verfügung stehen. Aber was auch super ist, dass es halt auch offene Vorlesungen gibt, komplette Vorlesungsreihen. Und da gibt es auch richtig, also welche mit, finde ich, toller Qualität hier von Leadmind hatten die noch lange auf ihrer Website jetzt ist das alles so nach youtube gewandert und wird ja alles zu google jetzt aber da gibt es tolle vorlesungsreihen dann

A1

Das sind wir jetzt eher so im über Machine Learning allgemein lernen, aber nicht über konkrete Modelle oder also schon so ein bisschen über konkrete Methoden vielleicht. Gibt es ein Introduction to Unsupervised Learning und dann lernt man da in 45 Minuten so ein bisschen die Übersicht, aber wenn du so nach konkreten Sachen suchst, mehr als dass du jetzt dir einen generellen Überblick verschaffen willst, fallen dir da noch Sachen ein?

I13

Boah, das dann ist es, da bin ich eigentlich immer direkt auf den Websites drauf. Also auf den Scikit-Learn, TensorFlow, schau mir da die Beispiele an. Ja, aber ich weiß auch gar nicht, zum Beispiel Variational Autoencoder. Ich muss einmal die Theorie verstehen, was macht der? Da gibt es unterschiedliche Abstraktionstiefen, bis mathematische Beschreibungen mit allen Details, so als ein Beispiel. Und dann weiß ich ja, wie das Ding funktioniert, dann will ich das bauen. Und da fehlt mir dann, wie baue ich das denn? Wie kriege ich das umgesetzt? Und dann hilft mir ja bei TensorFlow ein Beispiel, darauf setze ich auf und dann lese ich halt mehr und mehr über die einzelnen Layer, oder wie ich das verschalte, oder welche Funktion ich in die Klasse bauen muss. Also so gehe ich dann halt vor. Deshalb, eigentlich bin ich dann immer direkt schon bei dem, also einerseits gibt es dann wirklich die theoretische Beschreibung in den Vorlesungen und dann das praktische Anwenden aus der Toolbox mit der Doku, die da ist. So bin ich auch vorgegangen.

A1

Okay, also ist ja cool. Hintergrund der Frage ist, dass wir natürlich uns damit auch so ein bisschen, natürlich vergleichen wir uns auch mit anderen Systemen, die schon am Markt sind. Die Sachen, die uns jetzt eingefallen sind, da fehlen tatsächlich Vorlesungen, aber das geht natürlich in Richtung wissenschaftliche Publikation, weil die meisten Videos oder so, die sowas vorstellen, basieren ja dann irgendwie auch wieder, also so ein bisschen Wissenschaftskommunikation eigentlich, wissenschaftliche Ergebnisse irgendwie nutzbarer machen. Dann Library Documentations, Punkt 5 hier in der Liste, natürlich irgendwie auch schon viel diskutiert. Was noch ein bisschen näher dran ist an dem, was wir so gemacht haben, wären die Model Cards, ursprünglich von Google, aber gibt es jetzt zum Beispiel auch auf Hugging Face für so pre-trained Sprachmodelle oder auch die bildgenerativen Modelle und so. Ähnlich auch die Fact Sheets von IBM, die allerdings auch nur für IBM-Produkte irgendwie veröffentlicht werden und so on demand praktisch und nicht nur on demand, sondern auch on financial payment irgendwie bereitgestellt werden. Und Papers with Code, falls du es nicht kennst, vielleicht auch eine hilfreiche Ressource für dich nochmal. Jetzt für MobileNet V3 hier links das konkrete Paper, irgendwie 40 Seiten, kann man sich mal durchlesen oder halt auch nicht, wenn man die Zeit nicht hat. Rechts die Model Card, so wie sie auf Hugging Face steht, da sieht man dann zum Beispiel auch hier für Devices so ein paar Benchmark-Ergebnisse, Modellgröße und so ein bisschen so die Interface. Was für Bilder verarbeitet das Ding und so. Dann haben wir hier links auf Papers with Code, das ist ganz cool, alle möglichen Papiere und auch die dazugehörigen Codebases, in denen MobileNet in irgendeiner Form genutzt wurde oder vielleicht auch nochmal neu implementiert wurde und so. Da hat man auch für Datensätze dann so eine nette Timeline, wo man sieht, welche Modelle wie gut sind und so. Also für so Überblick-Sachen ist das auch ganz hilfreich. Rechts noch ein Blogpost, hier die Dokumentation aus Keras und noch ein generisches IBM Fact Sheet. Weil MobileNet wurde halt nicht von IBM veröffentlicht. Wenn man sich die ganzen Sachen so im Vergleich anschaut, kannst du da irgendwie Vorteile, Nachteile irgendwie mal nennen? Was fällt dir auf, was sind deine Gedanken dazu?

I13

Ja, man denkt dann wirklich direkt, wie schnell sehe ich was auf einen Blick. Da ist wirklich dein Label, glaube ich, am stärksten abstrahiert, am meisten eingedampft, am wenigsten Text und Bilder und man sieht es dann recht schnell. Ja, Paper, typisches Paper. Denkt mal, muss man viel lesen, viel Zeit investieren, um es zu verstehen. Dann das da drüber, das war glaube ich dieses Factsheet in der Mitte oben.

A1

Ja, ich glaube, das war die Modelcard tatsächlich.

I13

Die Modelcard. Ja, ich finde, die geht schon so in Richtung, was du hast, also auch schon weniger Text, mehr Zahlen, mehr Übersicht, eine Tabelle, also auch schon weiter abstrahiert. Und unten links ist es ja auch schon ähnlich, Tabelle, Code. Ja, also allgemein würde ich sagen, man sieht, alle sind Beschreibungen von Modellen und es gibt jetzt hier weniger starke oder mehr starke Abstraktionsebenen, die alles eindampfen so als Übersicht, ja, oder das ist eigentlich das, was du gemacht hast, ja jetzt wiederhole ich mich eigentlich zeigt dann eigentlich das am stärksten eingedampft und für den Anwender vielleicht doch am freundlichsten, ne?

A1

Das ist halt die Frage. Für den Anwender aus deiner Perspektive wäre ja so eine technische, etwas tiefer gehende Dokumentation mit direkter Schnittstelle, Code, API, wie zu benutzen und so vielleicht hilfreicher. Aber das ist ja wiederum was, was man halt dann irgendwie der Management-Ebene nicht zeigen kann. Sondern die wollen ja eine abstraktere Darstellung der Ergebnisse. Und dazu kommt ja auch, also wir verstehen auch dieses Label natürlich nicht als Ablöse von den anderen Systemen, sondern das ist halt eine Ergänzung. Und das, wo wir das Gefühl haben, das fehlt im Moment. Also diese abstraktere Darstellung, damit man diese Technologie auch verständlicher, transparenter machen kann, vielleicht auch nutzbarer machen kann, weil man kommt dann ja über den QR-Code auch wieder zu tiefer gehenden Informationen. Für Leute, die halt von der Materie noch nicht so viel verstehen.

I13

Ja, genau. Eigentlich ist es das, was ich auch eingangs gesagt habe, stimmt. Gute Ergänzung auf jeden Fall. Es ist immer wieder von der Anwendung abhängig. Genau. Mein Chef, klar, der würde jetzt vielleicht wirklich die Gewichtung einstellen. Oder wenn ich wüsste, was mein Chef will, dann würde ich auf die Website gehen, würde die Gewichtung einstellen, Enter drücken und dann würde ich mir dein Label rausziehen wahrscheinlich, wenn ich so vorgehen würde, wenn es das gäbe. Und dann würde ich mit über den QR-Code vielleicht doch ganz gerne auf diese Website draufgehen, wo es auch den Code gibt und die Paper-Beschreibung, um ein bisschen mehr zu verstehen, was da so passiert. Und so würde ich dann vielleicht vorgehen. Es ist dann halt wirklich immer von dem Anwender ausgedacht, was braucht er, wie abstrakt oder wie detailliert muss es sein.

A1

Ja, cool, vielen Dank. Kommen wir so ein bisschen zum Endteil des Interviews, wo wir uns nochmal so ein bisschen über das Stichwort Vertrauenswürdigkeit irgendwie austauschen können. Hast du das Gefühl, so wie du es gerade beschrieben hast, so eine Online-Webseite, da geht man hin, lädt seine Ergebnisse hoch und dann kriegt man so eine interaktive Darstellung, wo man dann einstellen kann, wie man das priorisieren will und so. Hast du das Gefühl, wäre das anerkannt oder gibt es da Probleme mit Vertrauenswürdigkeit? Also kannst du dich da irgendwie reinversetzen oder dazu was sagen?

I13

Ja, okay. Also, ja, das ist, glaube ich, da wieder von Mensch zu Mensch unterschiedlich. Ich bin eigentlich immer so eine Person, die nicht so, die nicht misstraut. Ich bin eigentlich immer offen und so, aber eigentlich müsste man, ja, so im Zuge jetzt von Russland, Hacking und all sowas, dass da irgendwie verzerrte Sachen bei rumkommen oder dass man sich dann irgendwie Schadsoftware runterlädt oder sowas, das wäre dann schon vielleicht kritisch oder dass man irgendwie Unternehmensgeheimnisse preisgibt, also wenn ich die kritische Brille aufsetze, dann wäre sowas

A1

Wäre das ein Problem? Also wenn du jetzt sagst, du hast wirklich so ein Data Frame, irgendwie Pandas oder irgendwas, eine CSV einfach, da stehen einfach nur Messwerte drin, im Endeffekt so eine Tabelle, ohne dass da jetzt steht, für welchen Einsatzzweck, auf welcher Maschine oder was auch immer. Ja, glaubst du, das wäre für XXXX, also glaubst du, das wäre im Rahmen deiner Arbeit bei XXXX, wäre das okay, wenn du das einfach hochlädst und dir dann irgendwie so Labels zurückgeben lässt und dann kannst du das irgendwie mit deinen Chefs besprechen oder würden die sofort sagen, wo hast du das denn her? Hast du unsere Daten preisgegeben? Was ist passiert?

I13

Ja, also gut wär's nicht, so würde ich sagen. Ich kann jetzt hier auch bei XXXX manchmal, wir haben ja unseren Chatbot hier auch, eigentlich sollen wir den halt immer nutzen. Manchmal sind die Antworten nicht ganz so gut, dann mache ich auch mal einen anderen, dann abstrahiere ich das Problem. Aber gut wär's nicht, also es wäre schon toleriert, wenn man sagen kann, gut, klar, ich brauchte das Labeling, das bringt uns ja auch nach vorne.

A1

gut ihr könnt die Software theoretisch noch lokal laufen lassen weil das Zeug was ich mache ist ja open source also ihr könntet halt die Software einfach lokal auf dem Rechner laufen lassen und müsste das nicht müsste die CSV nicht rausgeben das wäre vielleicht noch

I13

Ja das ist dann unbedenklich.

A1

ja gut so relativ du weißt ja nicht ob meine Software das im Hintergründe nicht wieder irgendwie an mich weiter schickt also das ist ein bisschen das Problem also du hast ja nicht die zeit die bei jedem open source Tool anzuschauen was jeder jetzt code Zeile macht, klar du kannst es irgendwie in einer abgesicherten VM ohne Internet Anbindung und so laufen lassen aber macht man das in der Praxis ist wieder

I13

Ja, ich meine, man muss ja immer ein Stück weit ein bisschen vertrauen. Wir vertrauen auch als XXXX Microsoft, wir sind eine Microsoft Company, die anderen sind eine Google Company. Hängt da vielleicht davon ab, wenn das jetzt wirklich von der XX kommt, dann weiß man, das kommt nicht aus Russland oder so. Das ist dann vielleicht okay. Ich sage halt, bei XXXX ist man nicht ganz so streng. Also jetzt ist man, ist die IT strenger geworden, die in der Entwicklung waren nicht ganz so streng. Ich durfte mir auch, bevor es jetzt wirklich starke IT-Einschränkungen gab, auch jede beliebige Python-Toolbox runterladen, da gab es keine Einschränkungen. Aber jetzt wirklich in den letzten Jahren ist das alles eingedampft worden. Das ist schon gar nicht mehr alles so einfach.

A1

Das ist auch nachvollziehbar, ne? Also Microsoft, wir hatten das ja jetzt erst vor ein paar Tagen, diesen absoluten Global Failure irgendwie mit einer Library, die auch von Microsoft als so semi-sicher eingestuft wurde und Microsoft selber sind jetzt auch nicht die Experten, was IT-Sicherheit angeht irgendwie. Also ja, klar, benutzen alle und unsere Uni wechselt auch immer mehr auf Microsoft, aber da gibt es halt auch krasse Sachen so.

I13

also hier bei mir, ich wurde jetzt einmal gehackt, nämlich in meiner Cloud hier und dann kam wirklich das war jetzt im November oder Dezember und das war hier meine Microsoft Cloud also Azure, genau, auf meinem Handy habe ich das immer Azure Storage vor allem Gott sei Dank hatte ich auch keine Daten drin aber auf einmal ging ich rein und hatte einen neuen Ordner der auf Kyrillisch Dokumente hieß und da hat er einen Monat später in die Nachricht, Microsoft wurde gehackt also, ja, aber ich sage halt ganz sicher ist man sich nicht, also das ist schon, hängt dann auch immer von den Daten ab, wenn man dann auf einmal vielleicht Business-Modelle machen will, die dann Business-Daten beinhalten, die dann wirklich Unternehmensergebnisse haben, ist das dann vielleicht schon kritischer, bei uns, ja, wird nicht gern gesehen, kann man aber mal machen in der Entwicklung

A1

Wenn man so über das Szenario, über das wir gerade gesprochen haben, da wären es ja keine wichtigen Secrets. Das wären ja halt nur so Measurements, wie schnell ist das Modell auf den Daten, du musst aber nicht die Daten rausgeben, du musst auch nicht mal beschreiben, was da gemacht wurde, das ist einfach irgendeine Berechnung, die passiert ist. Du reportest ja nur das, nur die sozusagen diese die Ergebnisse sozusagen dieser Berechnung, aber nicht die echten Datenergebnisse, sondern halt so Metriken über die Ergebnisse eigentlich. Und also da wäre wahrscheinlich sicherheitskritisch, da wäre jetzt nichts drin, wo man in irgendeiner Form Profit rausziehen könnte, was dann wiederum bei XXXX Schaden verursacht.

I13

Aber ich verstehe es schon, die CSV beinhaltet ja, das sind ja meine Input daten.

A1

Ne, du musst nicht die Input-Daten hochladen. Du lädst praktisch nur die CSV hoch, wo steht Modell XY auf Name des Datensatzes hat die und die Messwerte. Das heißt, du schickst keine Daten, die durch das Modell gehen hin und her, sondern nur Daten über das Verhalten deiner Masse.

I13

Ja, verstehe. Aber dann frage ich mich auch wieder, wenn XXXX jetzt ein Modell entwickelt, für ein Business Case, also nicht immer im technischen Bereich, ich stelle mir das gerade vor, im Accounting, im Controlling irgendwas, dann machen die das für ihren Datensatz und ich weiß jetzt gar nicht, ob die sich die Mühe machen würden, einen bekannten Datensatz zu nehmen, der ähnlich ist, der wie MNIST irgendwie oder ein anderer bekannter ist, den man nimmt und den da durchjagt, damit man eine Vergleichbarkeit hat. Also das würde es ja

A1

Wird’s ja vielleicht auch gar nicht geben. Aber, also, übrigens, wir laufen ein bisschen außer Zeit. Für mich ist das kein Problem. Aber falls du einen Follow-up-Termin hast...

I13

Ein bisschen können wir noch machen. bis 10 nach maximal.

A1

Ja, alles klar, kriegen wir hin. Ich wollte es nur einmal gesagt haben, genau. Also ich glaube, das Szenario wäre eher, du hast einen Datensatz, den du irgendwie aus welchem Business-Use-Case bei euch irgendwie raushast, dann hast du da 10 Modelle drauf ausprobiert und du willst praktisch reporten, welches Modell war wie gut und willst dann der Ebene drüber, der Management-Ebene, die Entscheidung überlassen, welches Modell dann produktiv am Ende eingesetzt werden soll. Da muss ja dann kein externer Datensatz eingeholt werden oder so, sondern du würdest den Code entwickeln für Modelle testen, das ist ja was, was du auch schon gemacht hast, machst du Hyperparameter-Suche, verschiedene Konfigurationen, testest das alles durch und findest dann dadurch raus, okay, das sind die Modelle, die wir getestet haben, die haben die und die Eigenschaften, die und die Vergleichswerte und das willst du jetzt verständlicher kommunizieren. Ich glaube, das wäre eher so der Use-Case.

I13

Wie mache ich das dann? Also dann habe ich zum Beispiel eine Software, die laden wir runter, die ist dann bei XXXX und dann schicke ich die Modelle dann da rein und muss aber auch meinen Input-Datensatz, mit dem ich das mache.

A1

Nö. Du kannst praktisch, du würdest Code haben, der macht einmal dieses, alle Experimente durchführen und die Messwerte als DataFrame oder CSV oder so rausschreiben und dann kannst du das wiederum an meine Software geben und die erstellt dann einfach nur diese Labels und speichert dir als PDFs. Also es ist einfach nur so eine Reporting-Ding, also es transformiert das, was du als Software rausschreibst in etwas, was du visualisieren kannst. Tatsächlich mit dieser Software, die ich geschrieben habe, hast du auch so eine interaktive Anwendung. Dann kannst du zum Beispiel sagen, ich will jetzt PowerDraw versus RunningTime vergleichen oder PowerDraw versus Accuracy, ist wahrscheinlich spannender. Dann hast du so einen schönen 2D-Scatterplot und kannst über jedes, über jeden Punkt, über jedes Modell hovern und kriegst dann das dazugehörige Label angezeigt und so. Also das, also ist ja jetzt nicht so, dass ich jetzt sage, ihr müsst das machen aber das ist so ein bisschen das, woran ich gearbeitet habe, was man halt machen könnte. So.

I13

Aber was ich jetzt nicht verstehe, ist zum Beispiel die Laufzeit. Wenn du sagst, ich muss nur die Ergebnisse

A1

Genau, die Laufzeit müsstest du halt auch messen und müsstest du halt irgendwie auch profilen dann in deinem Code.

I13

Aha, also gut. Okay, da gibt es so gewisse Schnittstellen. Ich muss gucken, dass zum Beispiel beim Output die und die Features mit getrackt werden, diese Metriken. Und wenn diese Spalten vorhanden sind, die kann ich dann bei dir hochladen und die wertet dann deine Software aus. Ja, verstehe ich.

A1

Genau, richtig. Also im Endeffekt habe ich, also ich habe das ja auch schon gemacht, also auch der Code, um das für ImageNet zum Beispiel zu machen, da kann man sich ja von inspirieren lassen und gucken, wie hat denn der XXXXXXX XXXXXXX das gemacht, als er das alles durchgemessen hat und das dann ähnlich nochmal selber macht und so. Und ich habe auch schon im Hinterkopf, ich habe vor, dass ich das, dass ich auch so eine Art Experimentier- API habe, wo man halt sagen kann, führe den und den Code aus und reporte die und die Metriken und so, dass man auch das als Service noch ein bisschen mehr irgendwie anbindet. Da ist man dann halt sehr schnell auch in so RapidMiner oder, also es gibt, also verschiedene Firmen haben auch an sowas schon gearbeitet, um so ein AI-Testbett(?) praktisch bereitzustellen.

I13

Ich sehe schon wirklich ja, der Aufwand, ich weiß nicht ob… Zum einen ist das natürlich schön dem Chef das zu zeigen, aber ich glaube meinem Chef würde schon reichen: Ach XXXXX cool dein Modell… Nh wenn ich 10 Modelle teste mit Hyperparametern und ich zeige ihm dann einfach anhand meiner Metrik wie gut das ist. Ich glaube das würde ihm ausreichen

A1

Also wenn die Qualität passt, dann ist der Rest auch okay.

I13

Ja, ich könnte mir nur vorstellen, und das ist jetzt wieder, ich weiß, da fehlt mir die Erfahrung, wenn ich in so einer richtigen Data Science Abteilung bin, die auch noch recht viel entwickelt, ob dann da vielleicht der Chef lieber sowas haben wollen würde, weil wir so ein bisschen… Mein Chef legt jetzt nicht allzu viel Wert auf Prozesse, er will das Ergebnis sehen, dass das läuft. Und fertig. Und wie das dann dargestellt ist, also wenn ich schon meine Metriken habe, dann reicht ihm das eigentlich

A1

Ja, Funktionalität ist immer das oberste Gut. Solange es funktioniert, ist der Rest dann nicht mehr so wichtig. Es ist wahrscheinlich auch, weil ihr arbeitet ja nicht an KI-Produkten, die ihr verkauft, in dem Sinne, dass ihr die Modelle bereitstellt, sondern ihr arbeitet, wenn ihr KI-Produkte macht, dann heißt das, das ist am Ende ein Gerät, was jemand kaufen kann oder ein Service, aber der kommt selber nicht in die Benutzung der KI sozusagen, sondern es wird halt alles eingebettet in ein anderes System. Aber ja, also um das an der Stelle dann hier ein bisschen aufzuräumen, auf jeden Fall vielen lieben Dank für die Einsichten. Super interessant und auf jeden Fall sehr coole Antworten auch drin. Zum Abschluss, ich werde dir die Folien hier zuschicken, falls du dir das nochmal in Ruhe anschauen willst, falls du da irgendwas rausgenommen hast. Falls dir jetzt noch irgendwelche Leute, ich meine, wir haben jetzt natürlich mit dir und mit XXXXX auch schon ein Interview gehabt, aber falls dir jetzt trotzdem noch Leute aus dem Bekanntenkreis oder auch bei anderen Firmen oder bei XXXX einfallen, die dir ein Video einfallen, wo du sagst, boah, ich habe mit dem da mal drüber gesprochen und ich glaube, der hätte auch noch spannende Dinge zuzusagen oder so, dann bitte nicht zögern. Immer gerne weiterleiten, die Webseite, die ist hier auch verlinkt und hilft uns auf jeden Fall. Wir haben jetzt irgendwie so zehn Interviews durch, wir wollen so 15 circa machen. Also wir sind gut dabei, aber ja, falls dir da Leute einfallen, immer gerne forwarden. Die Studie wird wahrscheinlich so bis Ende des Jahres schon dauern, grob, aber sobald wir Ergebnisse haben, kriegst du die dann auch nochmal per Mail. Genau. Aufnahme können wir, glaube ich, beenden.