Mein Name ist Elisabeth Knigge, ich bin 17 Jahre alt, komme aus Leipzig und beschäftige mich mittlerweile seit 2 Jahren mit KI und deren Potenzial. Vor allem starke künstliche Intelligenz, also KI, die den gleichen Grad an Intelligenz hat wie ein Mensch, interessiert mich dabei besonders. Deswegen versuche ich mit meinem Projekt einen kleinen Schritt in diese Richtung zu machen. Bevor ich jedoch mein Projekt vorstelle, möchte ich Ihnen zunächst eine Frage stellen: Was ist der Unterschied zwischen den zwei folgenden Aufnahmen?

Aufnahme 1

Aufnahme 2

Wie sie wahrscheinlich schon richtig erkannt haben, ist die Person in der ersten Aufnahme glücklich, während die Person in der zweiten Aufnahme traurig ist. Dies haben sie ganz unbewusst und wahrscheinlich ohne Aufwand feststellen können, obwohl der Inhalt des Satzes genau gleich ist. Für Computer ist es viel schwerer beides zu unterscheiden und einer Emotion zuzuordnen. Doch gerade diese Fähigkeit wird in den nächsten Jahren wahrscheinlich immer wichtiger werden. Denn mehr und mehr alltägliche Aufgaben, die auch stark mit Menschen zusammenhängen, werden von Robotern übernommen. Beispielsweise im Pflegesektor herrscht ein starker Personalmangel und schon jetzt gibt es erste Roboter Prototypen, die mit Patienten interagieren. Auch in der Telekommunikation spielen künstliche Assistenten eine immer größere Rolle. Ein großes Problem bei diesen Prototypen ist jedoch, dass sie menschliche Emotionen außerhalb von Gesichtsausdrücken meist nicht erkennen können und auch nicht in der Lage sind, selber Emotionen wie beispielsweise Mitgefühl zu zeigen. Um dieses Problem anzugehen, habe ich INGE programmiert eine KI, die aus englischen Sprachaufnahmen Emotionen klassifizieren kann.

Dafür habe ich ein Convolutional Neural Network kurz CNN mit der Keras-Bibliothek programmiert. CNNs benutzen in den Convolutional Layers eine mathematische Funktion, die nur die wichtigsten Datenpunkte eines Datensatzes zum Lernen übergibt, indem ein Raster über die Daten läuft und mit den Daten multipliziert wird, was man auch Convolution bzw auf Deutsch Faltung nennt. Gekoppelt mit so einem Convolutional Layer ist zweimal noch ein Pooling Layer, welches die Dimension der Matrix stark verkleinert, indem ein bestimmter Wert, wie beispielsweise der Maximalwert aus der ursprünglichen Matrix entnommen wird und einer neuen hinzugefügt wird. Diese Technik ermöglicht es dem Netzwerk, mit jedem Hidden Layer zunehmend komplexere Merkmale zu lernen. Damit das Netz aber überhaupt trainiert werden kann, müssen die Daten ins richtige Format gebracht werden, da Computer Töne nicht so verarbeiten wie wir. Dafür habe ich die Librosa-Bibliothek benutzt, die es ermöglicht Audio-Dateien zu laden, analysieren und zu bearbeiten. In Librosa gibt es ein Modul, dass Audiodateien in MFCCs umwandelt. MFCCs heißen ausgeschrieben Mel Frequency Cepstral Coefficients und übermitteln die wichtigsten Eigenschaften einer Audioaufnahme mithilfe von Matrizen. Jede dieser Matrizen enthält dann nicht nur eine kompakte Darstellung der Frequenzspektren der Aufnahme, sondern auch Metadaten wie Korrelation, die das CNN zum Trainieren verwenden kann.

Schlussendlich gibt INGE eine Matrix aus, die 8 Zahlen enthält. Jeder dieser Zahlen ist eine Klasse zugeordnet: So übergibt beispielsweise die 3. Zahl die Wahrscheinlichkeit dafür, dass die zugehörige Aufnahme die Emotion ‘glücklich” enthält. Insgesamt gibt es 8 Emotionen, die meine KI erkennen kann: neutral, glücklich, sauer, ruhig, angewidert, ängstlich, überrascht und traurig.

Beim Trainieren meines Models bin ich jedoch auf ein Problem gestoßen: Overfitting. Da die Daten so spezifisch waren, konnte INGE keine neuen Aufnahmen verarbeiten, da sie nur einen bestimmten Satztyp gelernt hatte. Deswegen trainiert die aktualisierte Version meines Projekts mit 3 Datensätzen, die ich zusammengefügt und normalisiert habe. Dadurch stieg die Flexibilität meiner KI deutlich. Mittlerweile erreicht sie auf dem Trainingsset eine Genauigkeit von ungefähr 70% und auf dem Testset eine Genauigkeit von 60%.

Zusammenfassend kann ich sagen, dass meine KI noch an einigen Punkten verbessert und erweitert werden kann. Dies sehe ich jedoch nicht als Kritikpunkt, sondern eher als Möglichkeit. So plane ich beispielsweise noch weitere Sprachen zu den Datensätzen hinzuzufügen. Mein größter Plan für INGE ist jedoch ein anderer: Ich möchte ihre gelernten Parameter dazu verwenden, eine weitere KI zu programmieren, die selber Emotionen sprachlich ausdrücken kann. Nach deren Fertigstellung, will ich beide so koppeln, dass sie sich immer weiter gegenseitig verbessern. Da INGE noch so viel Zukunftspotenzial hat, hoffe ich ins Finale zu kommen, um andere Menschen von meiner Idee zu überzeugen und INGE danach auch im größeren Stil weiter zu verbessern und zu erweitern.