**Annotated Bibliography CS488: Age detection using neural networks and machine learning**

**Pitch 1**

Neural networks is computing systems where machines learn from data and based on found patterns, machines make decisions and derive conclusions. Neural networks use interconnected nodes in a layered structure that depicts how the human brain works. Companies like Google use neural networks and machine learning to identify the right person to pass through security or predict somebody's reaction to a situation. For the capstone project, I plan to write a program using neural networks and machine learning concepts to guess a person's age from their provided photo to the program. Part of the linked dataset in the bottom will be used to train the program with identifying age, and part of it will be used for testing accuracy.

Dataset: <https://www.kaggle.com/datasets/arashnic/faces-age-detection-dataset>

**Annotated Bibliography for CS488 (1)**

Othmani, Alice. *Age Estimation From Faces Using Deep Learning: A Comparative Analysis*.

Science Direct, 2020,www.sciencedirect.com/science/article/abs/pii/S1077314220300424.

*Computer Vision and Image Understanding* is a research book and one of the chapters in the book is *Age estimation from faces using deep learning*. The paper has proposed many different Convolutional Neural Networks (CNN) algorithms based on their performance in detecting age. In the datasets that are used for the experiment the robustness of the best deep estimator is evaluated under noise, expression changes, different ethnicities, and different genders. In the research paper’s experiments the results demonstrate the high performance of the popular Convolutional Neural Networks frameworks against the state-of-art methods of automatic age estimation. A layer-wise transfer learning evaluation is done to study the optimal number of layers to fine-tune on Automatic Age Estimation (AAE). The Computer Vision and Image Understanding is related to my project because I’ll be looking at image analysis through convolutional neural network where I’ll also be working on finding the optimal number of layers to produce the best result in AAE.

**Annotated Bibliography for CS488 (2)**

Verma, Ishita. “Age Prediction Using Image Dataset Using Machine Learning.” https://www.ijitee.org/wp-content/uploads/papers/v8i12s3/L102010812S319.pdf.

In the cited source an unsupervised Convolution Neural Networks (CNNs) has been used for image processing and age detections. A CNN that handles multitasking, such as facial detection and emotional classification, is made by combining CNN and other algorithms and approaches. In the research, the object's face is first detected, then extracted from the photo, meaning the photo's background is eliminated first. The research has used 3 CNN layers and two layers with one final output layer. My research is very close to how *Age Prediction Using Image Dataset Using Machine Learning* research has approached age detection. I plan to use between two to three layers of convolution to get the closest result for predicting someone's age.

**Annotated Bibliography for CS488 (3)**

Singla, Sonia. “Age And Gender Detection Using Deep Learning.” *Analytics Vidhya*, 29 July

2022, www.analyticsvidhya.com/blog/2021/07/age-and-gender-detection-using-deep-learning.

*Age and Gender Detection Using Deep Learning* is a research paper that has used deep learning to predict the age of a person from photo. The dataset used in the research is comprised of age, gender, image, and pixels that are stored in .csv format and analyzed in Python programming language. Using Keras, an open-source Neural Networks library, images are processed to detect age and gender. After reading the dataset's data, photos' dimensions are converted to 48 by 48 to standardize images and sharpen the precision of age detection. The concepts and methods that are used in the study are somehow close to how I want to conduct my research. I'll use open-source python libraries on my dataset, then I’ll standardize the size of photos in the dataset, and I’ll analyze images based on their pixels and characteristics.

**Annotated Bibliography for CS488 (4)**

Zaghbani, Soumaya, et al. “Age estimation using deep learning.” ScienceDirect, May 2018,

[www.sciencedirect.com/science/article/abs/pii/S0045790617334298](http://www.sciencedirect.com/science/article/abs/pii/S0045790617334298).

In the cited work, human facial features such as posture, vocabulary, wrinkles, and intonation are seen as elements of facial analysis and age estimation. The research presents facial image age estimation based on autoencoders which is an artificial neural network used for unsupervised learning. One of the age estimation methods used in the study is by autoencoders, where a network of artificial neurons learns a hidden representation to reconstruct its inputs. For the purpose of my project, I find it interesting to use autoencoders and unsupervised learning, where the algorithm keeps learning from any image input even after training it with a test dataset. I’ll be using some of the concepts that are presented in the cited source and improve .

**Annotated Bibliography for CS488 (5)**

Chen, Shixing. Deep Age Estimation: From Classification to Ranking.

par.nsf.gov/servlets/purl/10074227. Accessed 27 Sept. 2022.

*Deep Age Estimation: From Classification to Ranking* is a research paper that has studied methods of human face analysis and age prediction. The images in the dataset have been categorized and each different approach of CNN has been trained with ordinal age labels. The study has used two methods of CNN: the first one is pre-training the program with

facial images and then fine-tuning with age-labeled faces. In the photo analysis, many features are studied, such as the distance of eyes from mouth to train AI with features that are easily noticeable to humans. The paper is very well organized in modeling how photos are analyzed

and how the algorithm works; it’ll be helpful for my paper to learn from the above cited source.

**Annotated Bibliography for CS488 (6)**

Qiu, Jiayan. *Convolutional Neural Network Based Age Estimation From Facial Image and*

*Depth Prediction From Single Image*. Jan. 2016, openresearch-

repository.anu.edu.au/bitstream/1885/102510/1/Qiu%20Thesis%202016.pdf.

The above cited thesis paper has used Convolutional Neural Networks (CNN) to predict age based on a given image to a program. In the thesis, multiple layers of CNN are used to analyze an image, and after analysis, through the filters, a prediction of age is given to the user. When an image Is given to the program first, a 4X4 patch of the image is used for analysis which is then down-sampled to a 2X2 patch by putting through the max pooling layer. This way, the center of the image, which most often shows a person's face, is used for age analysis and prediction. Based on how the cited thesis is structured, it would be helpful for my research if some ideas are implemented, such as max pooling. Methods such as max pooling will give my project more precision in predicting a person's age from a provided photo.

**Annotated Bibliography for DS488: Credit Card fraudulent detection using data analysis**

**Pitch 2**

Millions of credit cards have been stolen or hacked worldwide in the past few years. Most hacked or stolen credit cards have been used for fraudulent or illegal activities. Given that credit card fraud is a pressing issue in most societies, I’ll use Data Science Capstone Project as an opportunity to study the likelihood of a credit card being used for fraudulent activities. I’ll be working on modeling data to find the best fit that’s appropriate for Credit Card Fraud Detection. I’ll also look at what does the trend of credit cards being used for fraudulent activities look like in the future? Is there another variable that strongly correlates with the credit cards being stolen variable?

Dataset: <https://www.kaggle.com/code/hellbuoy/credit-card-fraud-detection/data>

**Annotated Bibliography for DS488 (1)**

*Ayorinde, K. (2021). A methodology for detecting credit card fraud [Master’s thesis, Minnesota State University, Mankato]. Cornerstone: A Collection of Scholarly and Creative Works for Minnesota State University, Mankato*. https://cornerstone.lib.mnsu.edu/etds/1168

Ayorinde’s thesis looks at the common trends in credit card fraud in the banking, retail, financial services, and healthcare industries. Credit cards that are stolen, misplaced, synthetic fraud, data breaches, mail interception, skimming, and merchant collision are common examples of ways credit cards are hacked or used for fraudulent activates. The thesis has used datasets and experiments with machine learning and models to classify fraudulent transactions. At the end of the thesis, the writer has researched the accuracy of the correct prediction based on the total number of input samples. My project is very much close to how Ayorinde has written *A methodology for detecting credit card fraud* thesis. I’ll use some of the exact concepts and methods that are used in the research paper such as Random Forest Decision Tree algorithm, and I’ll build layers of analysis on top to complete my research on credit card fraudulent detection. <https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?article=2167&context=etds>

**Annotated Bibliography for DS488 (2)**

“Data Science Project - Detect Credit Card Fraud with Machine Learning in R.” DataFlair, 17 July 2019, <https://data-flair.training/blogs/data-science-machine-learning-project-credit-card-fraud-detection/>.

In *Detect Credit Card Fraud with Machine Learning in R* blog, many algorithms are used in R programming language to detect credit card fraudulent activities. The program has used algorithms and methods such as Decision Trees, Logistic Regression, Artificial Neural Networks, and Gradient Boosting Classifier for data analysis. The blog has used the Card Transactions dataset, which contains a mix of fraud and non-fraudulent transactions in order to have an unbiased result. Sometimes some datasets can be unstructured, and for that purpose, the blog has used scale in data to standardize the dataset. Alongside data analysis, the blog has also built models such as logistic regression. The Detect Credit Card Fraud with Machine Learning blog is close to how I want to implement fraudulent credit card detection in my database project. I’ll be using a scale to standardize data, and I’ll also be building models for visual analysis, such as logistic regression.

**Annotated Bibliography for DS488 (3)**

Sharma, Nishant. *CREDIT CARD FRAUD DETECTION PREDICTIVE MODELING*. 2019, library.ndsu.edu/ir/bitstream/handle/10365/31611/Credit%20Card%20Fraud%20Detection%2 Predictive%20Modeling.pdf?sequence=1&isAllowed=y.

*Credit Card Fraud Detection Predictive Modeling* is a thesis paper that has applied data normalization as part of credit card fraud detection. For the experiment, the thesis paper used a dataset of transactions made by European cardholders during July 2015 through various credit cards. Before data analysis the data in the dataset is scaled so that the overall dataset is ready for a standardized modeling and analysis. Part of the analysis for the thesis is done using the Random Forest Decision Algorithm for classification and regression. Random Forest Decision Algorithm is also used to correct the habit of overfitting training sets, and a subset of the training set is randomly sampled to train each single decision tree, and finally a decision tree is built. The research paper is a good example of how I want to conduct my research through categorizing data and feeding different branches of the Random Forest Decision Algorithms with different parts of dataset and finally producing a predictive detection of credit card fraud.

**Annotated Bibliography for DS488 (4)**

Sulaiman, Bin Rejwan. “Review of Machine Learning Approach on Credit Card Fraud

Detection.” *SpringerLink*, 5 May 2022, link.springer.com/article/10.1007/s44230-022-00004-

0?error=cookies\_not\_supported&code=cd491904-38d5-4250-8864-803b33dfd7f5.

In *Review of Machine Learning Approach on Credit Card Fraud Detection*, machine learning is seen as an effective way of determining which transactions might be fraudulent and which are not. In the same way, as many other resources I’ve cited, Sulaiman has also used the Random Forest algorithm, which is a machine learning method that’s constructed from a decision tree. In the research paper, the Random Forest decision tree is seen as a slow algorithm in real time fraud analysis. The artificial Neural Network (ANN) Method is seen as a convenient algorithm for being an unsupervised method for predicting fraudulence. ANN and RFA find patterns in detecting fraudulent activities in credit cards, but ANN is believed to be an effective solution that I have to experiment with and learn. Since the research paper has analyzed credit card fraudulent activities with two different algorithms, it’s also helpful to me to experiment and learn different algorithms for my research paper.

**Annotated Bibliography for DS488 (5)**

Mekterović, Igor. “Credit Card Fraud Detection in Card-Not-Present Transactions: Where to

Invest?” *MDPI*, July 2021, [www.mdpi.com/2076-3417/11/15/6766/htm](http://www.mdpi.com/2076-3417/11/15/6766/htm).

The above-cited research paper is about detecting fraudulent activities of credit cards when cards are not present during transactions. The research has taken two approaches to predict fraudulent activities of credit cards; the first approach is supervised data analysis, and the second approach is unsupervised data analysis. For the dataset, the research has used 197,471 transactions from the past three from the research data. The transactions in the dataset are divided chronologically to achieve a realistic scenario when an actual credit card fraudulent activity happens. Ultimately, the research paper aims to find a cost-efficient way to detect credit card fraud. The dataset has also been modeled visually, where it’s easier to see the data flow. Based on the cited research, my research paper might be close to how the data is structured, and referring to the cited data might help me to structure my paper better.

**Annotated Bibliography for DS488 (6)**

Kristen, Barkved. “Credit Card Fraud Detection and AI”. [www.obviously.ai/post/credit-card-fraud-detection-with-machine-learning. Accessed 27 Sept. 2022](http://www.obviously.ai/post/credit-card-fraud-detection-with-machine-learning.%20Accessed%2027%20Sept.%202022).

*Credit Card Fraud Detection With Machine Learning* is an online blog introducing the concept

of credit card fraud and the types of fraud. The blog analysis why the traditional credit card

fraud detection methods don't work and why machine learning is a way to reduce credit card

fraudulence. The blog has stated that greater accuracy, less manual work, fewer declines in

normal transactions, and the ability to adapt and evolve machine learning elements have

helped reduce credit card fraud. Besides, the blog has brought examples of facts and data

about credit card fraudulence and how we can use that information to predict fraudulence. I

found the blog to be a good source to learn about machine learning and why the traditional

way of detecting credit card fraudulence didn't work, which can help me better write my

project.