**Success prediction in IT studies based on handwriting samples**

**Introduction**

This project represents the convergence of digital technology with traditional educational methodologies. It involves the digitization and analysis of historical academic records, highlighting a growing trend in educational innovation. Utilizing advanced image processing, machine learning and deep learning techniques, the project aims to transform paper-based archives into valuable digital datasets. This endeavor is reflective of a broader movement towards integrating technology into educational research and data management. Overall, it exemplifies the ongoing efforts to modernize educational practices through technological advancements.

**Necessary Background**

This field requires theoretical and practical knowledge about deep learning, machine learning image processing, natural language processing, and computer vision. We will use the following tools, languages, and libraries in the scope of the capstone

* **Python General Knowledge**
  + NumPy
  + Pandas
  + SciPy
  + Matplotlib
  + Seaborn
  + OpenCV
  + Python's Natural Language Toolkit (NLTK)
* **Working with Deep Learning libraries** 
  + PyTorch
  + TensorFlow
* **Traditional Machine Learning libraries**
  + Scikit-learn

**Literature Review**

Here you can find all the necessary articles that can be used as a good starting point for the capstone.

* Jubair, M. I., & Banik, P. (2012). A simplified method for handwritten character recognition from document image. International Journal of Computer Applications, 51(14).
* Liu, C. L., Nakashima, K., Sako, H., & Fujisawa, H. (2003). Handwritten digit recognition: benchmarking of state-of-the-art techniques. Pattern recognition, 36(10), 2271-2285.
* Ignat, A., & Aciobanitei, B. (2016, September). Handwritten digit recognition using rotations. In 2016 18th International symposium on symbolic and numeric algorithms for scientific computing (SYNASC) (pp. 303-306). IEEE.
* Ebrahimzadeh, R., & Jampour, M. (2014). Efficient handwritten digit recognition based on histogram of oriented gradients and SVM. International Journal of Computer Applications, 104(9).
* Vaidya, R., Trivedi, D., Satra, S., & Pimpale, M. (2018, April). Handwritten character recognition using deep-learning. In 2018 second international conference on inventive communication and computational technologies (ICICCT) (pp. 772-775). IEEE.
* Zin, T. T., Thant, S., Pwint, M. Z., & Ogino, T. (2021). Handwritten character recognition on android for basic education using convolutional neural network. Electronics, 10(8), 904.
* Dhanawade, A., Drode, A., Johnson, G., Rao, A., & Upadhya, S. (2020, March). Open CV based information extraction from cheques. In 2020 fourth international conference on computing methodologies and communication (ICCMC) (pp. 93-97). IEEE.
* Pal, U., Wakabayashi, T., & Kimura, F. (2009, July). Comparative study of Devnagari handwritten character recognition using different feature and classifiers. In 2009 10th International Conference on Document Analysis and Recognition (pp. 1111-1115). IEEE.
* Pradeep, J., Srinivasan, E., & Himavathi, S. (2011, April). Diagonal based feature extraction for handwritten character recognition system using neural network. In 2011 3rd international conference on electronics computer technology (Vol. 4, pp. 364-368). IEEE.
* Shrivastava, A., Jaggi, I., Gupta, S., & Gupta, D. (2019, October). Handwritten digit recognition using machine learning: a review. In 2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC) (pp. 322-326). IEEE.
* Pashine, S., Dixit, R., & Kushwah, R. (2021). Handwritten digit recognition using machine and deep learning algorithms. arXiv preprint arXiv:2106.12614.

**Dataset**

In the scope of this capstone project, we will collect our dataset by digitizing anonymous archival examination papers of alumni from the American University of Armenia (AUA). The dataset will encompass images captured under a variety of external conditions, such as disparate levels of brightness, contrast, and varying scanning methodologies. This diverse data acquisition approach is intended to yield a more comprehensive and nuanced understanding of the dataset's characteristics. Furthermore, this methodology is anticipated to enhance the robustness and adaptability of the machine learning or deep learning algorithms utilized in the project. By exposing these computational models to a broad spectrum of data variations, it is expected that they will develop an ability to process and interpret more generalized datasets effectively.

**Workload**

The workload for this capstone project will require a commitment of approximately 10 hours per week. This time will be distributed across various tasks such as research, dataset collection, coding, and documentation. Additionally, weekly meetings with supervisors will be conducted to discuss progress, address challenges, and receive guidance, ensuring the project aligns with expectations and goals.

**Coding Part**

In the scope of the capstone project, a key component involves the deployment of advanced image processing and computer vision algorithms to extract suitable features from the digitized dataset of examination papers. This dataset, derived from the alumni examinations, exhibits a diverse array of handwriting styles, each with unique attributes. The challenge lies in developing sophisticated code that can accurately identify and extract these distinctive features.

To address this, we will implement a range of image processing techniques. These techniques are designed to analyze the nuances in handwriting, acknowledging that individual characteristics manifest differently in each script. The processing will involve, but not be limited to, edge detection, contrast adjustment, and pattern recognition, all critical in isolating and identifying key handwriting features.

Additionally, the project will leverage the potential of machine learning and deep learning frameworks. The primary objective here is to utilize the extracted features from our dataset to predict various academic attributes of students, such as their academic proficiency and other suitable metrics. It is crucial to note that these predictions aim to provide insights into student learning patterns and should not be construed as definitive assessments of their capabilities.

The coding for these tasks will predominantly use Python, a language renowned for its extensive libraries and frameworks that are particularly adept in handling machine learning and image processing tasks. Libraries such as OpenCV for computer vision tasks, PyTorch for deep learning applications, and SciKit-Learn for traditional machine learning algorithms will be integral to our coding strategy.

**Problem Statement**

The primary challenge addressed in the scope of this capstone project is the feature extraction from archival examination papers of alumni, which are characterized by a diverse range of handwriting styles. This task involves not only the conversion of these documents into a digital format under varying external conditions but also the intricate processing and analysis of the unique handwriting features present in each document.

The objective is to develop a sophisticated computational framework capable of identifying and extracting these distinct features using advanced image processing and computer vision techniques. Additionally, the project aims to employ machine learning and deep learning algorithms to predict various academic attributes based on the extracted data. This necessitates the creation of a robust, adaptable codebase, predominantly in Python, to handle the complexities of the dataset and ensure accurate analysis and predictions.

**Methodology**

* Data Collection
* Preprocessing
* Feature Extraction
* Machine Learning Model Development
* Model Evaluation and Prediction

**Deliverables**

* Dataset
* Codebase
* Trained Models
* Documentation
* Presentation