- Nombre del proyecto:
    - Optimización del Rendimiento de Multiplicación General de Matrices (GEMM) en CPU-GPU: Predicción y Análisis de Parámetros
- Miembros del equipo:
    - Glenn Lozano Tapia (20150480)
    - Ronaldo Tunque Cahui (20140755)
    - Héctor Sánchez Domínguez (20130495)
- Conjunto de datos a utilizar:
    - SGEMM GPU Kernel Performance
      Paredes,Enrique and Ballester-Ripoll,Rafael. (2018). SGEMM GPU kernel performance. UCI Machine Learning Repository. https://doi.org/10.24432/C5MK70.
- Objetivo del proyecto:
    - El objetivo del proyecto es emplear técnicas de aprendizaje automático para predecir el tiempo estimado necesario para llevar a cabo la multiplicación general de matrices (GEMM) en un sistema específico que involucra una CPU con GPU. Esta predicción se basará en la exploración de diferentes configuraciones de parámetros. Se analizarán 14 configuraciones distintas del procesador, incluyendo el tamaño del grupo de trabajo local, la estructura de la memoria local, el factor de *loop-unrolling* del kernel, así como los anchos de vector para cargar y almacenar, entre otros aspectos. Además de la predicción, el estudio también tiene como objetivo identificar las configuraciones de parámetros que ejercen un mayor impacto en el tiempo de ejecución de esta operación, lo que proporcionará valiosos conocimientos para optimizar el rendimiento del sistema.
- Artículos científicos relevantes:
    - Agrawal, S., Bansal, A., & Rathor, S. (2018). Prediction of sgemm gpu kernel performance using supervised and unsupervised machine learning techniques. 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-7. https://doi.org/10.1109/ICCCNT.2018.8494023
    - Carvalho, P., Clua, E., Paes, A., Bentes, C., Lopes, B., & Drummond, L. M. d. A. (2020). Using machine learning techniques to analyze the performance of concurrent kernel execution on gpus. Future Generation Computer Systems, 113, 528-540. https://doi.org/10.1016/j.future.2020.07.038
    - Li, J., Ye, H., Tian, S., Li, X., & Zhang, J. (2022). A fine-grained prefetching scheme for dgemm kernels on gpu with auto-tuning compatibility. 2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 863-874. https://doi.org/10.1109/IPDPS53621.2022.00089
    - Matsumoto, K., Nakasato, N., Sakai, T., Yahagi, H., & Sedukhin, S. G. (2011). Multi-level optimization of matrix multiplication for gpu-equipped systems. Procedia Computer Science, 4, 342-351. https://doi.org/10.1016/j.procs.2011.04.036
    - Volkov, V., & Demmel, J. W. (2008). Benchmarking gpus to tune dense linear algebra. SC'08: Proceedings of the 2008 ACM/IEEE conference on Supercomputing, 1–11. https://doi.org/10.1109/SC.2008.5214359
- Propuesta tentativa de modelos de regresión a utilizar:
    - Regresión Lineal
    - Regresión Polinómica
    - Regresión con Bosques Aleatorios
    - Regresión con Incremento de Gradiente (por ejemplo, XGBoost, LightGBM)
- Declaración de trabajo grupal (según formato de la Directiva y normas para la elaboración de trabajos grupales)
    - *Adjunto en Paideia*.