Summer Internship 2023 IC/UNICAMP

- Objectives, areas of interest and schedule of each position are described below

- **Remuneração proposta: (2.000,00 – 3.000,00)**

Vantagens adicionais:
Vale Refeição ou Alimentação, Plano de Saúde, Plano odontológico, Transporte, Portal de vendas produtos Samsung, plataforma de orientação psicológica online.

- **Descrição do programa:** The candidate will work with our team of researchers, data scientists and health domain specialists in order to design and deliver innovative health and well-being features to Samsung devices and services. The candidate will work on projects that affect the life of millions of people worldwide, helping them to improve their health and wellbeing.
- **Resumo das vagas:** por projeto

_____

Project 1

Total number of positions: 2

Position 1:

Proposal: AI development intern

Objectives: Acquiring on-hands experience in deploying AI solutions in low-memory wearable devices and converting machine learning models between high level and low level languages. It is expected that the candidate develop skills and knowledge in the direct implementation and testing of machine learning methods in C language, as well as have the opportunity to study the different steps related to AI development and deployment such as Python → C transpilation.

Areas of interest:

- Machine learning

- C Language development and optimization

Grad level: Master or Doctorate

Schedule:
Week 1:
- Presentation of the company, the team, and the project
- Presenting the problems and goals related to the project
- Presenting the tools and frameworks employed by the team
- Creating and preparing the work environment

Week 2:
- Presentation of the main efforts (sub-projects) in the project (C code delivery and data preparation)
- Code review and test
- Presentation of the Python → C transpilation efforts in project

Weeks 3-11:
- Assisting on implementation and optimization of a Python → C translation for model used in project
- Code review and test

Week 12:
- Assisting on implementation and optimization of a Python → C translation for model used in project
- Code review and test
- Presenting the final results to the team

Position 2:

Proposal: Data scientist intern

Objectives: Acquiring on-hands experience in Healthcare Data Science leading to predictive models in low-memory wearable devices. It is expected that the candidate develop skills and knowledge in medical data processing, visualization and quality assessment.

Areas of interest:

- Data science

- Health science

Grad level: Master or Doctorate

Schedule:
Week 1:
- Presentation of the company, the team, and the project
- Presenting the problems and goals related to the project
- Presenting the tools and frameworks employed by the team
- Creating and preparing the work environment

Week 2:
- Presentation of the main efforts (sub-projects) in the project (C code delivery and Data preparation)
- Assisting on new database quality assessment and visualization

Weeks 3-11:
- Assisting on new database quality assessment and visualization

Week 12:
- Assisting on new database quality assessment and visualization
- Presenting the final results to the team

_____

Project 2

Total number of positions: 1

 Position 1:

Proposal: AI researcher intern

Objectives: The candidate will be part of our awesome team of engineers, data scientists, and researchers to build cutting edge technologies and implementing the latest research into our products impacting the life of millions of people worldwide.

Areas of interest:

- Machine learning

- Data Analysis

Grad level: Master or Doctorate

Schedule:

Week 1:
- Presenting the company, the team, and the project
- Presenting the problems and goals related to the project
- Intern software engineering training
- Creating and preparing the work environment

Week 2:
- Introducing the main concepts related to the project
- Introducing the tools and frameworks employed by the team
- Introducing the code and pipelines developed by the team
- Introducing the data to be used

Week 3:
- Literature review
- Defining the problem and goals

Weeks 4-5:
- Analysis of the available data
- Proposal of machine learning architecture

Weeks 5-9:
- Designing and code the models using our pipelines and frameworks
- Training, evaluating, and updating the developed models
- Comparing results with other known models
- Analyzing and improving the proposed model
- Reviewing the goals and proposed solutions

Weeks 10-12:
- Analysis and review of the results obtained so far
- Tests and code review
- Writing a scientific paper/technical report
- Writing documentation
- Presenting the final results to the team

_____

Project 3

Total number of positions: 3
Positions 1, 2, and 3:

Proposal: AI researcher intern (1), Data scientist intern (2), AI development intern (3)
Objectives: The candidate will be part of our awesome team of engineers, data scientists, and researchers to build cutting edge technologies and implementing the latest research into our products impacting the life of millions of people worldwide.
Areas of interests: The candidate will have the opportunity to work in one or more out of three areas of specialization:
- Model embedding
- Wearable sensors study
- Machine learning modeling
- Data analysis and processing

Grad level: Master or doctorate

Schedule:

Week 1:
- Presenting the company, the team, and the project
- Presenting the problems and goals related to the project
- Samsung software engineering training
- Creating and preparing the work environment

Week 2:
- Introducing the main concepts related to the project
- Introducing the tools and frameworks employed by the team
- Introducing the code and pipelines developed by the team
- Introducing the data to be used

Week 3:
- Literature review
- Defining the problem and goals

Weeks 4 and 5:
- (1) Researcher - Machine learning modeling
  - Time-series modeling exploration
  - Neural Architecture Search (NAS)

- (2) Data science - Wearable sensors study
  - Evaluate protocol data quality
  - Feature engineering

- (3) Developer - Model embedding
  - Model compression techniques
  - Upgrade data engineering pipelines
  - Transpiling (source-to-source)

Weeks 6:
- Study Presentation
- Designing and coding experiments
- Analyzing and improving the proposed approaches

Weeks 7, 8 & 9:
- Reviewing the goals and proposed solutions
- Exploring other ways to solve the problem
- Comparing results with other known approaches
- Integrating proposed solution into framework

Weeks 10, 11 & 12:
- Writing a technical report /scientific paper
- Writing documentation

- Presenting the final results to the team

_____

Project 4

Total number of positions: 2

<u>**Position 1:**</u>

Proposal: Data Scientist/Development intern

Objectives: The candidate will work on a health related solution proposed to Samsung's devices that combines data collection, processing and evaluation based on machine learning techniques. The candidate will face a problem related to integration of project solutions to organize and aggregate the results in the different prediction conditions that the project presents. The solutions involve implementing Python and C applications to achieve the automation of algorithm predictions. We expect the candidate to improve his/her skills and knowledge in the areas of data analysis and processing, to enhance the solution currently in development inside the team.

Areas of interest:
- Development
- Data science
- Health science

Grad. level: Master or Doctorate

Schedule:

Week 1:
- Presenting the company, the team, the project, environments and platforms
- Presenting the problems and goals related to the project
- Presenting the tools and frameworks employed by the team
- Creating and preparing the work environment

Weeks 2 and 3:
- Introducing the main concepts related to the project
- Introducing the code and pipelines developed by the team
- Introducing the machine learning algorithms developed by the team
- Introducing the data used to train the models

Weeks 4-9:
- Defining the problem and goals
- Understand the different types of results needed to integrate into the final results aggregation
- Understand the expected output formats and discuss the best format

- Implement the application using Python and C

Weeks 10, 11 and 12:
- Analysis and review of the results obtained so far
- Tests and code review
- Writing a scientific paper/technical report
- Writing documentation to consult
- Presenting the final results to the team

Position 2:

Proposal: Data Scientist/Development intern

Objectives: The candidate will work on a health and well-being related solution proposed to Samsung's devices that combines data collection, processing, analysis and evaluation based on machine learning techniques. The candidate will face a problem related to data visualization and develop a Python application or program to improve the current pipeline step. We expect the candidate to improve his/her skills and knowledge in the areas of development, data processing, validation, aggregation, analysis and data storytelling (dashboard).

Areas of interest:
- Data science
- Development
- Data visualization

Grad. level: Master or Doctorate

Schedule:
  Week 1:
- Presenting the company, the team, the project, environments and platforms
- Presenting the problems and goals related to the project (general overview)
- Presenting the tools and frameworks employed by the team
- Creating and preparing the work environment
  Weeks 2 and 3:
- Introducing the main concepts related to the project
- Introducing the data used by the team
- Introducing the data processing pipeline developed by the team
  Weeks 4, 5 and 6:
- Defining the problem and goals (data synchronization)
- Review the literature related to synchronization between signals

- Getting in touch with the data, its nuances and configuration
- Data cleaning and preparation to deal with the tasks
- Framework and tools necessary to development (dashboards)

Weeks 7, 8, and 9:
- Dashboard development applying learned techniques from previous week
- Incorporate changes on current data processing framework

Weeks 10, 11 and 12:
- Analysis and review of the results obtained so far
- Tests and code review
- Writing a technical report
- Writing documentation to consult
- Presenting the final results to the team

_____

Project 5

Total number of positions: 2
Positions 1 and 2:
Proposal: AI researcher intern
Objectives: Explore machine learning models for noise canceling in biomedical signals. The candidate is expected to develop skills and knowledge related to implementing and evaluating machine learning models for time series and will have the opportunity to study different models and adapt state-of-the-art solutions.

Areas of interest:
- Machine Learning
- Data Science
- Health Science
- Development (Python and C)

Grad level: Master or Doctorate.

Schedule:
Week 1:
- Presentation of the company, the team, and the project
- Presenting the problems and goals related to the project
- Presenting the tools and frameworks employed by the team
- Creating and preparing the work environment

Week 2:

- Presentation of biomedical signals and its physiological meaning
- Presentation of the datasets
- Presentation of the efforts related to removing signal noise
- Set development environment
- Explore classic signal processing algorithms in Python and C

Week 3:

Study of classic solutions for noise-canceling

Study implementation of adaptive filters in Python and C

Explore the dataset of biomedical signals

Definition of a baseline solution

Weeks 4 and 5:

Literature review of machine learning models for noise-canceling

Implement and evaluate literature solutions for noise-canceling

Define evaluation metrics

Evaluate baseline solution

Weeks 6, 7, 8, 9, and 10:

- Adapt literature solutions
- Propose new solutions
- Explore generative models

Weeks 11 and 12:

- Write a technical/scientific report about the findings
- Present findings to the team

_____

Project 6

Total number of positions: 3

Positions 1, 2, and 3:

Proposal: AI researcher intern (1), Data scientist intern (2), AI development intern (3)

Objectives: Explore machine learning models for health related solution proposed to Samsung's devices. The candidate will work on tasks that combines data collection, processing and evaluation based on machine learning techniques. Acquiring on-hands experience in deploying AI solutions in low-memory wearable devices and converting machine learning models between high level and low level languages.

Areas of interests: The candidate will have the opportunity to work in one or more out of three areas of specialization:

- Model embedding
- Machine learning modeling

- Data analysis and processing

Grad level: Master or doctorate

Schedule:
  Week 1:
- Presenting the company, the team, and the project
- Presenting the problems and goals related to the project
- Samsung software engineering training
- Creating and preparing the work environment

  Week 2:
- Introducing the main concepts related to the project
- Introducing the tools and frameworks employed by the team
- Introducing the code and pipelines developed by the team
- Introducing the data to be used

  Week 3:
- Literature review
- Defining the problem and goals

  Weeks 4 and 5:
- (1) Researcher - Machine learning modeling
    - Time-series modeling exploration
    - Neural Architecture Search (NAS)
    - Data augmentation

- (2) Data science - Wearable sensors study
    - Evaluate protocol data quality
    - Feature engineering
    - Upgrade data engineering pipelines

- (3) Developer - Model embedding
    - Model compression and hardware acceleration techniques
    - Transpiling (source-to-source)
    - Upgrade data engineering pipelines

  Weeks 6:
- Study Presentation
- Designing and coding experiments
- Analyzing and improving the proposed approaches

  Weeks 7, 8 & 9:
- Reviewing the goals and proposed solutions

- Exploring other ways to solve the problem
- Comparing results with other known approaches
- Integrating proposed solution into framework

Weeks 10, 11 & 12:
- Writing a technical report /scientific paper
- Writing documentation
- Presenting the final results to the team

_____

Project 7

Total number of positions: 1

<u>Position 1:</u>

Proposal: Data scientist intern (1)

Objectives: Explore machine learning models and analyze health related data produced by Samsung mobile devices. The candidate will work on tasks that involve signals acquisition and processing as well as implementing machine learning models for time series predictions.

Areas of interest: The candidate will have the opportunity to work in one or more of the following areas:

- Machine learning modeling

- Data analysis and processing

Grad level: Master or Doctorate

Schedule:

Week 1-2:
- Presentation of the company, the team, and the project
- Presentation of the problems and goals related to the project
- Presentation of the tools and frameworks employed by the team
- Presentation of the results obtained so far by the team
- Creating and preparing the work environment
- Presentation of the main efforts (sub-projects) in the project
- Code reviewing and testing

Weeks 3-9:
- Assisting on analyzing and validating data produced by data collection protocols
- Studying wearable sensors and evaluate protocol data quality
- Perform feature engineering

- Implementation of solutions to test and analyze data
- Code optimization
- Code review and test

Weeks 10-12:
- Assisting on implementation and optimization of a Python model used in project
- Code reviewing and testing
- Writing documentation
- Writing a technical report /scientific paper
- Presenting the final results to the team