This exercise is consisted of two parts. **Part A** revolves around designing a solution that bridges the EU-funded project objectives with EXUS Financial Suite banking SW. **Part B** revolves around building a prediction module for a particular case study.

Part A

EXERCISE GOALS

The scope of the exercise is to design a solution addressing the needs of a concept elaborated in the context of an EU-funded project (Section A) taking into account and covering as much as possible a set of requirements (Section B) stemming from the EFS product (EXUS Financial Suite banking SW). In specific, the candidate shall:

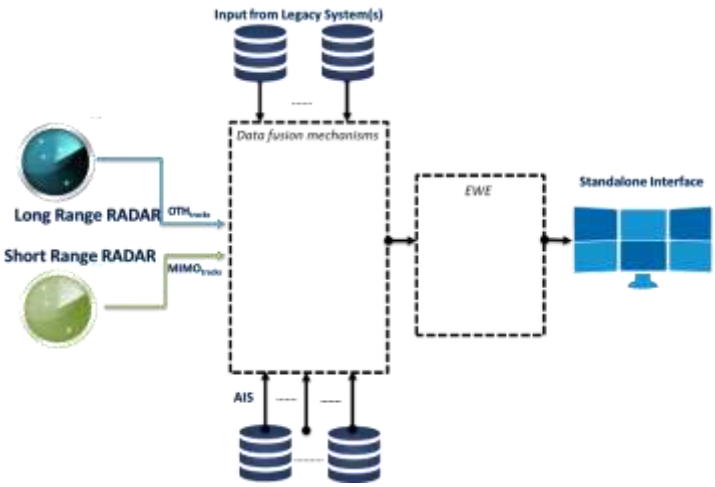
a) **Propose a high-level architecture that solve the PROJECT's goals (Section A) taking into account the data integration and data flow of all available data sources, combined so we can answer to the requirements of both Project and EFS (as many as possible)**
b) **Based on the high-level architecture, propose an Artificial intelligence / Machine learning framework-approach answering to the requirements of both Project and EFS (as many as possible) referring major components**
c) **Suggest recent technology trends that could support the aforementioned architecture**

Deliverables for Part A

1. You should submit **a report** (in pdf – 1500 words) where you present your design and your proposals justifying your choices.

SECTION A – PROJECT CONCEPT

PROJECT will apply a modular and scalable architecture that will exploit the outputs of the its sensors along with data collected from existing surveillance systems in order to enhance detection, identification, and recognition accuracy and range.



The **PROJECT sensors** (a long-distance radar and a supportive short-range high resolution radar) will be modular blocks within the PROJECT architecture that will deliver each a **list of tracks**. Their physical sites will be connected to the backend using an internet connection, through a dedicated satellite link if the sites are isolated. It is assumed that sensor tracks will go through a pre-filtering at sensor level. The list of tracks through the PROJECT sensors are then being further filtered and fused within a **Data fusion mechanism** that also incorporates information (when available) from existing surveillance or other legacy systems (such as AIS).

The Data fusion mechanism will further improve the list of tracks generated (accuracy of position, direction, speed), while predicting the future state. Data are available both in batch or online mode (real-time data). The specific outputs will be further processed through an **Early Warning Engine (EWE)** that will have inherent knowledge related to local resources, list of past incidents and threat assessments that will enable a knowledge based heuristic decision making that will also be the final output of the PROJECT platform overall.

The outputs of the PROJECT platform will be graphically presented to a **standalone local User Interface**.

It should be noted that the PROJECT architecture will be designed to be both scalable and modular in terms of its components and outputs.

(Project Goals) The high level outcome of the proposed solution is recommendations on required interventions based on risk assessment and self-training of threat detection models. Potential steps (but not mandatory) that lead to the desired outcome are:

a) **A threat classification of all simultaneously detected targets based on AIS data, historical data in available databases as well as maneuvering patterns of detected and tracked vessels.**
b) **Automatic Target Recognition (ATR) through cross correlation of all input data.**
c) **Target Continuous Tracking, especially valuable for high-threat vessels.**
d) **Alarms including collision warning, boundary violation and proximity alerts.**

Glossary

1. AIS - The automatic identification system (AIS) is an automatic tracking system that provides information such as unique identification, position, course, and speed of vessels. Vessels can choose whether they transmit continuously or not these data.
2. Legacy systems are:
   a. Local short range radars (other than the two aforementioned radars) that provide data regarding position, speed and course of vessels
   b. Geospatial satellite data (provided once per 3 hours) indicating vessel position
   c. Weather data
3. Novel long and short-range radars developed within PROJECT are providing position, speed and course of vessels data

SECTION B - REQUIREMENTS STEMMING FROM EFS PRODUCT.

Data provided are:

Debt collection system database:

- **Client:** information about his age, marital status, occupation, products he owns e.g., debit cards
- **Products**: Description of the products
- **Cases**: each case corresponds to a specific client and a specific product it is updated with information about the debt that this client has for the specific product, e.g., due days, debt amount, actions made by the bank such as a call or send an SMS, *recording of the calls*, and the following actions made by the client.

Note: in the debt collection systems' database a client inserts only when he has a debt (there is due past amount).

Bank database:

- **Client:** information about his age, marital status, occupation, banking account he, products he owns e.g., debit cards
- **Account:** date of creation, statements, transactions etc.
- **Products**: Description of the products
- **Cases**: each case corresponds to a specific client and a specific product it is updated with information about made by the client e.g., amount he pays and date

Open data sources, probable open data sources which could be used in order to gather more useful information about the client as the following (make the assumption that you can have the needed information available without any restrictions)**:**

- **Social media**

- **Tax authorities services**

Note: Consider that the information coming from the open data sources is provided via REST API.

Required EFS output:

- **Classification of the clients of a bank with debt in different categories as Trusted, Not trusted etc.**
- **Alarm raising when a client changes trust status**
- **Recommendation of strategic approach of client in order to make debt collection more effective**

Glossary

1. EFS is a comprehensive suite of software applications that manages credit risk along the whole lifecycle of accounts, from the moment of disbursement until write-off or debt sale. Through EFS, the bank can organize and automate it's operations regarding collections and recoveries, from segmenting portfolios to applying relative strategies and managing contact channels. All communication with the customer like phone calls, SMS's or letters as long as external agencies management or legal processes operates under EFS control and all gathered info is used by EFS rule engine in order to decide the next steps of the process.

Part B

EXERCISE GOALS

The scope of the exercise is to implement a classifier for a case study.

Deliverables for Part B

1. You should submit at least one **Python script or Jupyter notebook**. Letting us see how and why you went about our challenge in a given way is integral to completing this task.

2. You should submit efficient, clean and sufficiently commented code. Your submission should be hosted on a **private git repository** and invite the reviewers (they will be reported through an e-mail) to your personal repository. Group the contents into the following sub-directories, omitting the empty ones.

- Notebooks /scripts
- data (only include new data files, if any, excluding the files we provide)
- artifacts
- code
- misc
- readme.md

3. You should submit **a presentation in PowerPoint** where you present in high level the design and the results of your solution. You should also submit **a report** (in pdf - 1000 words) describing the process followed and key results.

Case Study

One of the most simple and popular consumer finance products is an Unsecured Personal Loan, which is a loan that is issued and supported only by the borrower's creditworthiness, rather than by any type of collateral. Competition in the market is strong and margins are slim so the product has to be well prepared. To be able to offer the right price for a bank (i.e., EXUS' customer), has to know what is the cost of lending and a big part of that cost comes from non-performing loans. For every euro earned there will be a few pence lost and understanding loss rates is an important exercise which will help to inform pricing.

You will find four tables with random sample data from an external source. Since only 20% of the initial population failed to pay their debts fully, the sample was stratified to increase the share of accounts gone bad. The new ratio is 50:50 and therefore weights need to be adjusted. Records are of different ages and some loans could have been active at the cut-off date of 2019-10-01. It should be noticed that not all loan types are of interest. Terms below 12 and over 60 months are out of scope and not eligible. Also, data needs tidying up a bit.

Exploratory Data Analysis and Machine Learning Model Development

1. Explore the data in depth to get an understanding of the problem and how to proceed in the later steps.
2. Build a classifier that predicts probability of a customer being default and evaluate its performance given this dataset.
3. We suggest you focus on fundamentals and robust modelling – split strategy, avoiding data leakage, examining your metrics – rather than getting the best performance, yet expect your model to make some sense of the data. We leave trying extra steps and algorithms, if any, to you.
4. State all of the assumptions.

Relevant files in data directory:

- Default_Data.csv: data describing defaulters and no defaulters (i.e., good/bad customers).
- Loan_Information.csv: data providing information related to loans.
- Credit_Bureau_Information.csv: data providing risk-based information.
- Client_Information.csv: data providing additional information for the customers.

Default dataset dictionary:

- UID - customer ID
- Recordnumber - customer's loan id
- Dval - default value
- Dmon - month on books when default was registered

Loan information dataset dictionary:

- UID - customer ID
- Recordnumber - customer's loan id
- Openbalance - loan opening balance
- Accstartdate - account opening date
- First_month - month when customer first logged into online servicing
- Last_month - month when customer last logged into online servicing
- Repayperiod - original loan term
- Searchdate - date when customer data was uploaded to the database

Credit bureau information dataset dictionary:

- UID - customer ID
- Score - acquisition score
- Class - customer type

Client information dataset dictionary:

- Includes large number of features. Use exploratory data analysis approach to establish predictive features!

Tips

1. **Tell us a story!** We expect you to be creative and inquisitive. Imagine that you will present your results to a team of technical and non-technical employees.
2. **Show us your data skills.** Some of the data skills we look for are:
    a. cleaning, data preparation, exploration, visualization
    b. target selection (labelling, defining success metrics)
    c. modelling (e.g., tuning, explanation of model selection)
    d. code (clean, commented)
3. **Back up assumptions with data.** Assumptions can be made but should be stated and backed up with data where possible.
4. **Play to your strengths.** Spend more time on visualization, statistics, machine learning, or business insights depending on your skills.
5. **Keep it simple!**

Good luck!!!