Summer Training Conducted By:-

Dr. Mohamed Abbassy

Final Report Conducted By:-

Ali Mohamed Sayed Ahmed Ali

Trainee ID:-

20221449583

Program:-

Internal Deep Learning

Introduction:

In this course, you learn to use SAS Viya to create supervised machine-learning models. You need to go from data to decisions as quickly as possible. Machine learning models are at the heart of critical business decisions. They can identify new opportunities and enable you to manage uncertainty and risks.To create these models, you need powerful and easy-to-use software that can help you wrangle your data into shape and quickly create many accurate predictive models. You also need an integrated process to manage your analytical models for optimal performance throughout their lifespan.SAS Viya provides efficient, repeatable processes and a reliable architecture for managing data, communicating the rationale, and tracing the predictive analytics models through the deployment phase

Aspects of the training course: -

1. Apply the analytical life cycle to business need
2. Incorporate a business-problem-solving approach in daily activities
3. Prepare and explore data for analytical model development
    a. Data Exploration
    b. Feature Extraction
    c. Input Transformations
    d. Feature Selection
    e. Variable Clustering
    f. Selecting Your Algorithm
4. Create and select features for predictive modeling
5. Develop a series of supervised learning models based on different techniques such as decision tree, ensemble of trees (forest and gradient boosting), neural networks, and support vector machines.
6. Comparing models within a pipeline or across pipelines.
7. Evaluate and select the best model based on business needs
8. Deploy and manage analytical models under production.

ACKNOWLEDGEMENT

It is a pleasure to acknowledge many people who knowingly and unwittingly helped us, to complete our Training. First of all, let us thank God for all the blessings, which carried us through this training duration. Also, we want to thank here as our training supervisor Dr. **Mohamed Abbassy** who has always stood by our side and guided, appreciated and encouraged us. Continuing the same, he enlightened us in the various stages during the training and provided us with many insights and useful examples, which proved to be of immense help in successful completion of this training.

Training Name:

- Internal deep learning (SAS Viya)

The topics of the course:

-Machine Learning in Business Decision Making	-Introduction to SAS Viya	-Data Exploration
-Feature Extraction	-Input Transformations	-Feature Selection
-Selecting Algorithm	-Decision Tree	-Tree-Structure Models
-Recursive Partitioning	-Pruning	-Ensembles of Trees
-Neural Network	-Network Architecture	-Network Learning and Optimization
- Support Vector Machines	-Large-Margin Linear Classifier	-Methods of Solution
-Nonlinear Classifier: Kernel Trick	-Model Assessment and Comparison	-Model Deployment

Date and Period:

Training for the groups:

- The training course started on July.19.2023 and completed Augest.20.2022.
- The duration of the training is 60 hours for 4 Weeks.

Place of training:

- The activity took place at the hall (No.404) of Faculty of Computing and Data Science, Alexandria University.

The instructors:

- One trainer participated in the activity as follows:

1. Dr. Mohamed Abbassy	taught 60 hours
Total	**60 hours**

The target groups:

Around 150 representatives from Alexandria University participated in this activity. List of names attached.

Methodology of training:

The training focused on using practical methods such as :

- Brainstorming regards the definition of role in leading and developing the organization.
- Presenting the subjects by involving the trainees in discussion and encouraging dialogue.
- Dividing the large group into small groups.
- give these small groups specific tasks.

Subject of the course:

- **Data Science Tools**

  - Using impressive range of E-Learning resources from SAS .
  - Communicate effectively, accurately, and confidently through solid work, a focus on high-frequency data and regular well-Structured.
  - Job Roles.
  - Those needing to improve their data science skills, whether for business needs, educational advancement, or everyday conversation.
  - Prerequisites
  - Functions & Situations

- **Impression Management**

  - What is the management and importance
  - Effective presentation
  - Effective communication techniques.
  - Interpersonal communication.
  - Business communication.
  - Curriculum development.
  - Raising organizations' communication
  - Interpersonal impression management

- **Presentation / Branding**

  - Writing an effective report
  - Ability to translate complex concepts into digestible content.
  - Why is a training important
  - Fact sheet
  - Different types of tasks
  - Functional training
  - Presentation and Other information
  - What are employers looking for
  - Tips

- **Time management**

  - Tasks time management.
  - How to use time.
  - Create a daily planner.
  - Planning, allocating, setting goals, analysis of time spent, monitoring, organizing and scheduling.

- **Coaching and leadership**

  - Leadership skills.
  - Skills management.
  - Teams and groups.
  - Presentation skills.
  - Active listening.

- **Data science Skills**

  - Concepts of Information Technology (IT)
  - Presentation

The training details

- Arrival 19 July 2023: Presentation of participants, Getting-to know-each-other
    - 09:00 AM Arrival of participants
    - 09:30 AM welcome by host team.

- **First Week:**
    - Machine Learning in Business Decision Making
        - Creating a Project and Loading Data
    - Essentials of Supervised Prediction
        - Modifying the Data Partition
        - Building a Pipeline from a Basic Template
    - Introduction to SAS Viya
        - SAS Viya: A Deeper Dive
    - Data Preparation
        - Exploring Source Data
        - Modifying and Correcting Source Data
        - Alternate Method for Modifying and Correcting Source Data Using the Manage Variables Node
- Define Training Target "Customer Churn"
- Create a new project in Model Studio based on the commsdata data set.
- Modify metadata roles of some variables, explore the advanced project settings, and change the data partition properties
- Build a new pipeline from a basic template for class target.
- Data Exploration node in Model Studio to assay and explore a data source.

- **Second Week:**
    - Feature Extraction
        - Adding Text Mining Features
    - Input Transformations
    - Feature Selection
        - Selecting Features
        - Saving a Pipeline to the Exchange.
    - Variable Clustering
        - Clustering Inputs for Data Reduction
    - Selecting Your Algorithm
        - Logistic Regression
        - Linear Regression

- Create new features using the Text Mining node.
- Transformations node to apply a numerical transformation to input variables.
- Variable Selection node to reduce the number of inputs for modeling.
- Save the pipeline to the Exchange, where it is available for other users.
- Logistic regression for nominal and binary target.

- **Third Week:**
    - Decision Trees and Ensembles of Trees
    - Tree-Structure Models
    - Recursive Partitioning
    - Pruning
    - Ensembles of Trees
    - Network Architecture
    - Network Learning and Optimization
- Building a Neural Network Model
- Improving a Neural Network Model by Changing the Network Architecture Parameters
- Improving a Neural Network Model by Changing the Network Learning and Optimization Parameters
- Improving a Decision Tree Model by Changing the Tree Structure Parameters
- Improving a Decision Tree Model by Changing the Recursive Partitioning Parameters
- Improving a Decision Tree Model by Changing the Pruning Parameters

- **Fourth Week:**
    - Support Vector Machines
    - Support Vector Machines and Additional Topics
    - Large-Margin Linear Classifier
    - Nonlinear Classifier: Kernel Trick
    - Additional Tools
    - Model Assessment and Deployment
    - Model Comparison
    - Model Deployment
- Building a Support Vector Machine Based on Default Settings
- Changing the Methods of Solution for a Support Vector Machine
- Modify the kernel function and other parameters and compare this model performance to the other model in the pipeline.
- Compare the models' performances based on different fit statistics.
- Pipeline Comparison tab enables you to compare the best models from each pipeline created
- Exploring the Features for Scoring and Running a Scoring Test in Model Manager

Positive aspect of the training:

- The participants were interested enthusiastic about all the subject.
- The instruction followed a democratic system in managing their sessions.
- The participants have gained reasonable skills experiences which will enable them to prepare and write project project proposals.
- More exercises and tasks.
- The hall was comfortable and very convenient.
- Hospitality for the trainees was acceptable.
- The relationships among the participants were in harmony.

Method of training:

Using the following methods:

- Evaluation each meeting by participants.
- Brainstorming Methods
- Collective interactions.
- Discussion groups, mini.
- Practical exercises, individually and collectively.

Training Project:

Project Number	Dataset	Goal	Models Used	Team
Pro.No.6	HOUSEPRICE	Predict the value of houses in Boston	1- SVM 2- Linear Regression 3- Quantile Regression 4- Neural Network 5- Forest	1- Ali Mohamed Sayed 2- Abdullah Hussein Ibrahiem 3- Abdelrahman Ashraf Rageb 4- Fares Mohamed Fathy 5- Raghadan Ramdan Mohamed

Recommendations:

- The participants are hoping to get other training in Data Engineering with SAS.
- To explain decision Trees in more details.
- To get all needed martial in the first week of training.

Evaluation:

- Training has been evaluated by the participants through practice, evaluation, Project, and the results were very satisfactory. The rate of evaluation %50

Attachments:

1. Training Material.
2. Dataset Practice.
3. Data Description.
4. Certificate from SAS