DAIR Knowledge Package Development

Statement of Work

Name of Proposed Knowledge Package:
Machine Learning Model Generation and Validation for Time-Series Prediction

SME / Company (“Lead Contractor”): BluWave-ai

TABLE OF CONTENTS:

Information Section

About this pilot

CANARIE is undertaking a pilot project as part of the evolution of the Digital Accelerator for Innovation and Research (DAIR) program to provide Small and Medium Enterprises (SMEs) with the opportunity to package their expert knowledge of advanced technologies and share this expertise in the form of what we are calling Knowledge Packages.

Knowledge packages are supplied by SMEs who have developed specialized knowledge in applying a particular advanced technology in a vertical market segment. The packages will be curated by CANARIE in the form of a browsable catalog to be shared with the entire DAIR participant community. Knowledge packages are comprised of a concise documentation set and sample code (under an open source, free license) which may be deployed for study and characterization by any DAIR participant onto the DAIR cloud.

The goal of the Knowledge Package project is to reduce the time required by SMEs to research and evaluate advanced technologies they may be considering in their cloud products or services. Specifically, SMEs will share their knowledge with other SMEs in the form of: lessons learned through individual research, valued topical resources found on the web, best practices, and considerations or recommendations for others considering the technology for their own application. Benefits to SMEs producing and sharing their knowledge with others include: documenting developed knowledge, internal training material for new staff, recognition as an expert in the field, increased public visibility of you and your company, and giving back to the community of Canadian SMEs. Finally, sample code provided within a Knowledge Package must not contain proprietary intellectual property in whole or part and must be made available free of use by DAIR SMEs (public) in accordance with the DAIR Participation Agreement and Acceptable Use Policy.

Under this pilot, CANARIE will pay SMEs for the development, delivery and support of a Knowledge Package. The pilot period will begin no earlier than September 4, 2018 and end no later than March 30, 2020.

Introduction to Technology and Benefits of Use

Machine Learning Model Generation and Validation for Time-Series Prediction provides a deployable reference solution for DAIR participants to observe and study application of machine learning techniques in time-series prediction. DAIR participants interested in time-series prediction would find this content useful in their approach toward implementing their own predictor.

The application opportunity domains are broad with the following as examples:

- Energy optimization: AI-based smart grid optimization, particularly in the presence of highly variable and distributed renewables such as wind and solar.
- Smart City: smart traffic control, security detection and dispatch, real-time transportation optimization
- Border security: e.g. shipment threat detection and control
- Network and Systems security: intrusion detection and control
- Infrastructure management: sensor-based deterioration detection, maintenance scheduling, depreciation and renewal optimization
- Real-time logistics: item placement, location, loading, and shipping optimization
- Medical applications: predicting future health risks, or recovery prediction, based on time series health diagnostics information

What is the problem?
The primary objective of time-series prediction is to develop models that provide plausible future values of a time-series, given past historical observation of same and other time-series.

Why the problem is relevant (and to whom) and worth solving?

Time-series, i.e., a sequence of data taken at successive equal time-intervals, are prevalent in numerous applications in statistics, finance, meteorology, natural sciences, and engineering. Forecasting a time-series enables predictive actions that adjust system behaviour with respect to a plausible event in future. Example of time-series include electricity demand and price, fuel price, agricultural yield, stock market closing price, and renewable resources output.

What are the traditional or present methods of addressing the problem?

Conventionally, time-series prediction is performed through statistical inference techniques that take into account past observation of same time-series to forecast future values.

Why the solution description later in the document provides a compelling means of addressing the problem relative to traditional solutions?

Compared to conventional time-series analysis techniques, cutting edge-machine learning models offer several advantages, including the followings:

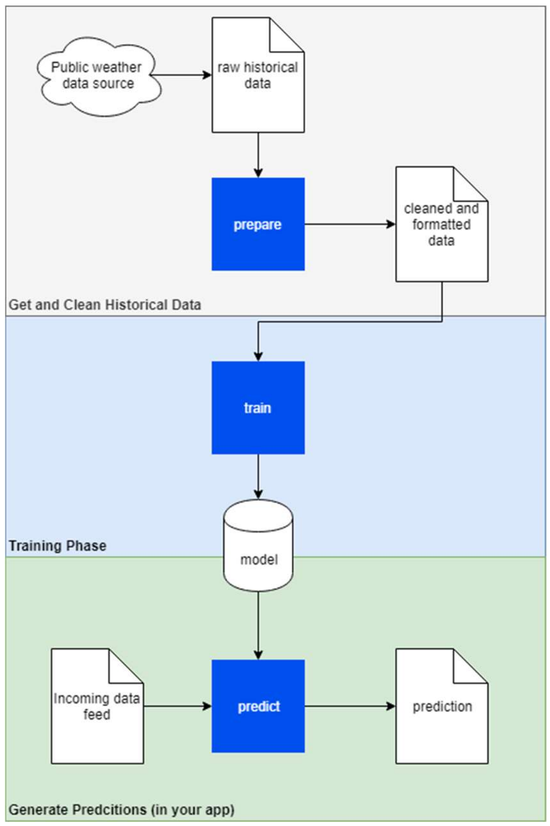
1. Handling larger amount of data: longer historical observation period can be ingested, as well as several parallel time-series that may contribute to the prediction accuracy of the target series.

2. Robust to data pre-processing: state-of-the-art machine learning techniques can handle raw data, while conventional statistical inference models are highly sensitive to bad historical data such as outliers or missing values.
3. Learning through a sequence: Models such as Recurrent Neural Networks (RNN) have attracted massive interest in time-series prediction due to their ability to exhibit temporal dynamic behavior.
4. Automation friendly: Machine learning techniques can be automated more effectively, offering continuous training and self improvement.

Solution Description

Solution Overview

This solution will train a wind generator output predictor using historical data including wind turbine past measurement and meteorological data that will be obtained from a public weather data source. We'll then walk through data processing with the "prepare" application, followed by training of our Machine Learning method in the "train" block. Within the "train" block, we'll explain how to gauge and score the prediction model to decide when to terminate the training process. Then we'll show how to use the trained model via the "predict" application that is paired with a simulated incoming data feed to produce the prediction.



Component Descriptions

Component	Description
Raw historical data	Data downloaded from a public weather source in its raw form (that we can't use as is with the trainer). Historical data from wind turbine measurement will be processed as well.
Prepare	An application that will prepare and convert the Raw historical data into the cleaned and formatted data the trainer is expecting. (Python)
Cleaned and formatted data	Data is cleaned, and in the format the trainer is expecting
Train	An application that will produce a model from a cleaned data (Python, Machine Learning, Tensorflow)
Model	The output from the trainer

Incoming data feed	Incoming data that we're making a prediction with
Predict	Application that uses the Model + Incoming data feed to make a prediction (Python, Machine Learning, Tensorflow)
Prediction	The expected next value based on the incoming data feed

Mandatory criteria:

You must address each of the following mandatory criteria:

Criteria	
1. The proposal is consistent with the Expression of Interest communicated and aligns with market directions for leading edge technology.	Yes
2. The proposed Knowledge Package will make use of DAIR cloud infrastructure (networking, compute, storage, cloud service(s)/APIs and/or sensors).	It will use of DAIR: • Networking • 2x Compute (4 core, 8GB RAM) o With 80GB of compute disk • Object Storage: o 20G object storage for bundled sample data set • 1x GPU (for optionally accelerated Tensorflow Training)
3. How will the proposed Knowledge Package contribute to one or more of <u>CANARIE's expected results:</u> (a) Enhance opportunities for collaborative knowledge exchange within Canada's SME community through the maintenance and development of the DAIR Knowledge Packages and related tools and services; (b) Expand Canadian SMEs access to and utilization of the DAIR Cloud and the availability of tools and programming that increase the effectiveness of its use; and, (c) Enable the creation of innovative Information and Communications Technology (ICT) products and	a) It will develop a machine learning model for time-series prediction that can be extended to several other application domains, in a way that is easy to understand, replicate, and play with. b) This KP will work out of the box with data that can be obtained from a public weather source now, and into the future. It gives a good example of Machine Learning for Time-Series prediction that DAIR makes easy to consume and learn from. c) Machine Learning Time-Series prediction is useful in a variety of domains and is the leading edge in a human efficient method for handling complex time-series that mutates over time.

services and accelerate their commercialization in Canada.	
4. The proposed Knowledge Package must not already exist as an open source project or affordable commercial equivalent that could be used in place.	Google searching has not yielded anything as complete, nor well packaged, nor easy to use to quickly start and explore as the proposed Knowledge pack. There are many references about the topic but they are disjointed, only discuss a small subset, and put the burden on the consumer for how to correctly search and pull together the right information to successfully build a time-series predictor.
5. Knowledge Package development must be technologically and economically viable within the timeline of the pilot.	Yes
6. Software developed under CANARIE funding **must** be made available for other SMEs to use at no cost through the CANARIE DAIR cloud program for a period of up to 5 years from the beginning of development, or until it is declared end-of life. Note that all software must fully support the CANARIE DAIR hybrid-cloud environment described in the following documents: Architecture: DAIR 2.0 Solution Architecture Integration: DAIR 2.0 Package Integration Specification	Yes
7. All funded work must be performed in Canada.	Yes
8. The Lead Contractor must be a Canadian small or medium-size company (SME < 500 employees).	Yes, we a SME < 15 employees https://bluwave-ai.com

Lead Contractor

Lead Contractor	
Organization Name	BluWave-ai Inc
Organization Type	Incorporated
Anticipated role/s in the project	Design, Produce, Support

Project Authority (PA) The person who will take direct responsibility for completion of the proposed project. This person must be associated with the Lead Contractor and will be the main point of contact.	
Name	Hubert Sugeng
Title	Director Engineering
Phone	613-454-5041
Email	hubert.sugeng@bluwave-ai.com

Project Manager (PM) CANARIE strongly recommends that your team include a dedicated Project Manager. If you are planning to include a Project Manager but have not yet hired/identified one, please type “Yes” into the Name field below. If you are not planning to include a Project Manager, please type “No” into the Name field below. The PA may also be the PM.	
Name	Hubert Sugeng
Title	Director Engineering
Phone	613-454-5041
Email	hubert.sugeng@bluwave-ai.com

Lead Software Engineer/Developer The person who will take direct responsibility for developing source code and/or documentation assets included in the proposed project. This person must be associated with the Lead Contractor.	
Name	Mostafa Farrokhabadi
Title	Senior Smartgrid/AI Scientist
Phone	
Email	mostafa.farrokhabadi@bluwave-ai.com

Mostafa Farrokhabadi is a Senior Smartgrid/AI Scientist at BluWave-ai. He has more than 6 years of experience in designing mission critical grid solutions for industry and academia. Part of his industrial and academic career has developed cutting-edge predictive control and optimization systems, including AI-based time-series predictors for grid components such as renewable energy resources. Mostafa obtained his PhD in Electrical and Computer Engineering from the University of Waterloo, during which he received multiple research and teaching awards. He has published in several peer-reviewed high-impact venues, and has obtained comprehensive program certificates in machine learning and data science. Mostafa has also studied and performed research in Sweden (KTH Royal Institute of Technology) and Germany (Karlsruhe Institute of Technology) and have been an invited speaker, presenter, and panelist at several renowned electrical engineering conferences around the world.

Potential Uses

Smart Grids

Smart grids are typified by the increased penetration of data-driven technology in various application domains. Advanced metering infrastructure is being deployed, gathering a large volume of high-resolution data. Thus, several players including electric utilities intend to harness the power of big data and enhance system predictive control. In addition, with the proliferation of renewables energy resources, on-line electric vehicles, and smart loads in active distribution networks, intelligent coordination of agents is necessary for stable and economical operation, as most of these components are stochastic in nature and exhibit intermittent performance. Thus, there is a compelling need for analytical methods to ingest large amount of data and provide predictive control and optimization actions that ensure system reliable and optimal operation. In this context, accurate time-series predictors are of paramount importance to mitigate uncertainties associated with the increased penetration of renewable energy resources, electric vehicles, etc.

Stock Market Predictor

Predict the next price (based on past data) of a Stock Market element. Given the right set of data relevant to the Stock Market element, a predictor could be made on the price of the index. This could be used for using a Machine Learning method to forecast the next expected price to either choose to buy, sell, or hold.

Future customization and/or extension of functionality

The reference software will give a generic structure to a machine learning project, how to download data from a public weather source, and how to process it with a Machine Learning Time-Series Predictor. Extending the functionality would involve replacing the data from a public weather data source with the SME's desired data set, understanding the problem domain, exploration of feature selection, and altering the Time-Series Predictor code to match.

Special Purpose Equipment

None

Project Plan

KEY TASKS LIST

	Key Task	Start Date yy/mm/dd	End Date yy/mm/dd	% of total effort required for the project	Milestones / Deliverables - must be tangible and measurable
1	Develop, and test	2019/02/04	2019/03/01	40%	Test Plan results
2	Integrate into DAIR	2019/03/04	2019/03/15	20%	DAIR cloud launch template produced
3	Write Reference Solution Description & Knowledge Kit	2019/03/11	2019/03/29	20%	Reference Solution Description available on Wordpress for approval. Knowledge Kit Template available on Wordpress for approval.
4	Solution Testing & Bug Fixes	2019/04/01	2019/04/12	20%	Delivered in DAIR Cloud

Risk Assessment and Mitigation Plan

Loss of staff resources. Mitigated by rehiring and by budgets incorporating this person on the payroll.

Software Provenance

Who would authorize software releases? The Project Manager

What validation procedures would be completed prior to release? A release test plan.

How would you deal with upgrades / patches to third party software packages that you might use?

We would shield ourselves from it by having the Knowledge Pack not require any updates, nor downloads of it to function.

Testing Plan

QA to validate the requirements with test plans written based on the requirements. A separate person that was not the designer, nor developer, is assigned the QA role to verifies it is usable and performs its intended functions by testing the Knowledge Pack on DAIR following the documents produced.

Intellectual Property

You acknowledge and agree that all right title and interest in DAIR and the intellectual property related thereto is and remains that of CANARIE and any enhancement or contribution to the intellectual property of CANARIE made as a result of the User's use of DAIR shall belong exclusively to CANARIE.

Appendices

None.