

**Using AI and machine learning for predictive analytics in estimating the growth in the demand of delivery services post COVID-19**

**Proposal**



**Submitted by**

**STUDENT NAME**

**STUDENT ID**

# ABSTRACT

In the present research project “Using AI and machine learning for predictive analytics in estimating the growth in the demand of delivery services post COVID-19” the System of Predictive Analytics(SPA) is proposed, the challenges on its creation are considered.

***Keywords***—AI, Cloud Computing, Machine Learning, Prediction of Time Series, predictive analytics, Concept Drift

Table of Contents

[1.1 Background (overview of topic and motivation) 1](#__RefHeading___Toc3558_3973182502)

[1.2 Aim 2](#__RefHeading___Toc3560_3973182502)

[1.3. Objectives 2](#__RefHeading___Toc3562_3973182502)

[1.4. Project Overview 3](#__RefHeading___Toc3564_3973182502)

[1.4.1. Scope (what will I do? How it work?) 3](#__RefHeading___Toc3566_3973182502)

[1.4.2. Audience (Who is it for?) 3](#__RefHeading___Toc3568_3973182502)

[2.1.Summary of existing approaches (e.g. Competitive analysis, if appropriate) 3](#__RefHeading___Toc3570_3973182502)

[2.2. Brief summary of existing literature (e.g. Annotated bibliography or initial literature review with brief summary of sources) 4](#__RefHeading___Toc3669_3973182502)

[3.1.Approach (Description of the research and the development methodology, e.g. software development model, requirement gathering model, test and evaluation process) 5](#__RefHeading___Toc3574_3973182502)

[3.2.Technology (implementation tools and resources, such hardware and software) 6](#__RefHeading___Toc3576_3973182502)

[3.3.Version Management plan (e.g. Git repository or shared drive) 6](#__RefHeading___Toc3578_3973182502)

[4.1.Tasks required to complete each objectives 7](#__RefHeading___Toc3580_3973182502)

[4.2.Schedule (i.e. Gannt or other, showing activities, deadlines) 7](#__RefHeading___Toc3582_3973182502)

[4.3.Data management plan (e.g. Google folders for project logs, reports, literature etc) 8](#__RefHeading___Toc3584_3973182502)

[4.4.Deliverables 8](#__RefHeading___Toc3586_3973182502)

# 1.Introduction

## Background (overview of topic and motivation)

Predictive analytics is one of the most important types of business data analytics, because it is aimed at making predictions about forecasting trends and behaviors. Thus correct and reliable predictive analytics allows making correct business solutions like ordering the cheapest accessories, produce the most demanding goods, predict the most essential services, etc. In order to be practical, the methods of this analytics should generate insights for the future with a controllable degree of precision.

Predictive analytics is used in a wide range of organizations, and its global market is currently approximately $10.95 billion by 2022, growing at a compound annual growth rate (CAGR) of around 21 percent in the years 2016 - 2022.

The predictions can be based on historical data and different techniques such as statistical modeling, numerical equation solution and others. Current boom of artificial intelligence techniques and rapid growth of computing services propels the trend of using machine learning and AI in predictive analytics. The using of ML and AI techniques not only helps in calculation and improving of precision for the well-known methods, like Regression, Clustering, etc., but also became the source for designing new approaches for analytics by building, for example, complex neuron networks that successfully predicts future sales, outcome, etc.

One of the challenges, that were faced before the whole world, businesses, and economies in the modern world, is the spreading of COVID-19, that has impacted the way of living of the masses and societies. The need of enterprises and markets to be resilient under the threat of the pandemic stimulate the managements to design the new models, that are capable to calculate the impact of COVID-19.

The reason to concentrate the efforts on the problem of demand of delivery services post COVID-19 is the fact that the pandemic affects this services deeply, that leads to the change of the situation on this market. However, some of the predictions that were produced at the start of the pandemic happened to be wrong. For example, some analytics had supposed that the demand on the food delivery services would rapidly rise during lock-downs, but that forecast wasn’t correct for all the regions and various factors affects the real growth of demand services. Therefore, the investigation of the factors that affects demand services sales and discovery of the reliable model of this growth in post COVID-19 world is the important and valuable task, that is demanding for good solution.

The upheaval of the past COBID-19 has raised up the imperative to use Concept Drift (CD) detection in ML applications. Thus, the goal of the project is to design the SW system that uses CD for improvement of a prediction accuracy in ML for delivery services.

## 1.2 Aim

The aim of the project is to design overall structure of the computer system, that allows to design, implement, test, and experimenting on ML models for predictions of the sales of the Demand Services (DS) from the historical data of the sales, based on the Time Series(TS) type of information. These data includes information about the kind of services and characteristics of the customers, transport and welfare in the area, so on. Also this system is capable for the user to modify these data and input some other crucial and hard for formalization parameters as the lock-down type, state of the lock-down in the area, etc.

## 1.3. Objectives

The objectives of the proposed project:

- to design the structure of the SPA systems and their interactions with the databases (DB) for their operation in post COVID-19 market environment, to present its dataflow;

- to propose the unit that allows to input the data non-related to DB for the changes in prognosis detection;

- to propose and implement the updates in the structure of data for time-series (TS) prediction with Concept Drift detection methods for SPA as the unit of the system that tweaks predictions methods for different condition

- to present recommendation for the SW of SPA implementation and testing.

## 1.4. Project Overview

### 1.4.1. Scope (what will I do? How it work?)

The designed SPA algorithms and system allows to operate with different methods of TS forecasting and analytical methods, like Autoregression, Moving Average, etc by using DB with historical data on the DS and information from Internet for more correct PA in times of change. The proposed SPA allows for the user (data analytic) to modify the data and input non database type of data for tweaking the performance of forecasting methods. Also, these models can be evaluated on the base of their correctness by various metrics on run-time data.

This system is also wielded with continuous and active learning SW means, that improve SPA performance by adjusting the parameters of the models on the run, an CD detection algorithms are helping to adapt these learning models for different kinds of lock-downs and pandemic situations in automated mode.

### 1.4.2. Audience (Who is it for?)

The proposed system is designed for the business analytics who are struggling to adapt their predictive models with the new post COVID-19 reality. The solution should help to create the new models and tweak the existing ones for the presence of the new data, that should be considered for evaluating the impact of COVID-19 on predictions of sales. Also, the proposed SPA SW methods are supposed to suppress this negative effects of existing approaches in predictive analysis.

# 2.Background review

## 2.1.Summary of existing approaches (e.g. Competitive analysis, if appropriate)

There are some examples of drift-aware systems with different algorithmic traits and use cases. For examples, **simple** and **smart retraining** systems – that just retrain models on detected CD, Learn++ that updates voting weights of each model on CD, Streaming Ensemble Algorithm (SEA), Dynamic Weighted Majority (DWM) / AddExp and some others.

These approaches have their advantages and shortcomings, and it is not clear which one of them is the most suitable for the considered task. However, the Error Intersection Approach showed a good result for the similar task, therefore we believe it can be regarded as a good starting point.

There are open SW frameworks, that could be used for testing (e.g. [*https://github.com/rsdevigo/pyAdwin*](https://github.com/rsdevigo/pyAdwin)). We select CD\_MOA(Bifet,2013) as the base implementations.

## 2.2. Brief summary of existing literature (e.g. Annotated bibliography or initial literature review with brief summary of sources)

There is a vast amount of works dedicated to Concept Drift detection and correction[], including survey papers, that cover versatile methods in ML field. For example, Zliobait ̇et al (2015), describe different mathematical approaches for CD and also their practical implementations and tests approaches. Many of them regarding Time Series predictions and Business analytics trends forecasting.

M. Straat et al. 2022, describe how CD evaluations can improve the work of the neural network, regression and clustering methods and provide the scheme of practical implementation of a framework that uses CD and predictive techniques simultaneously.

In the work of Lucas Baier et al.,2020, presented the method of Error Intersection Approach with practical examples that showed effectiveness of using CD detection for the task of business prediction combined with regression and ML techniques.

Bahrami Bidoni, Zeynab & Montreuil, Benoit, 2021, showed the practical example of the system, that can use ML models and evaluation of changes in data for correct predicting of behavior of urban logistics in the times of change.

# 3.Methodology

## 3.1.Approach (Description of the research and the development methodology, e.g. software development model, requirement gathering model, test and evaluation process)

Some research should be done to create optimal database structure, therefore planned to perform some investigation on open databases for data, and design the structure of the DB.

Next step is the Concept Drift algorithms implementation to detect rapid changes for the obtained data.

Then, algorithm that allows for Time Series to be tweaked in the concept drift detection is going to be implemented. The design of the tests for the system should be developed.

And, final testing, is process of the system design in the real-world production should be described.

## 3.2.Technology (implementation tools and resources, such hardware and software)

The list of required specialized resources in this proposed research work against the purpose is summarized in table 1.

|  |  |
| --- | --- |
| Required Resource | Usage |
| Access to Online Digital Libraries i.e. Google Scholar, IEEE Xplore, Science Direct, etc. | To collect the relevant research literature to perform an extensive review of literature pertinent to the proposed research area. |
| Cloud to store Database | Cloud to store Database |
| Github | Repository to store code |
| C++ | Programming language for the implementation |
| SQLite | Database for keeping data |
| Linux Ubuntu 20 | Open OS for working with Python and SQLite |
| PC | Computer for working |

Table 1: List of required specialized resources in proposed research work

## 3.3.Version Management plan (e.g. Git repository or shared drive)

The database will be stored on the local PC and Server. The backup copy of it must be created every week.

The code of the models stored in the Git repository on GitHub.

# 4. Project Management

## 4.1.Tasks required to complete each objectives

The development of the project consists of the next steps:

- Literature review;

- Database structure design;

- Flexible dataflow design for SPA system, that includes the information connected to the pandemic or the other important information for PA;

- Research of the databases for predictive analytics on delivery services;

- Implementation of the Concept Drift(CD) tweaking methods using C++;

- Recommendation of ML model implementation development;

- Developing of the recommendations for the testing and evaluating of the designed system, Testing coordination of DB unit and CD units.

## 4.2.Schedule (i.e. Gannt or other, showing activities, deadlines)

The of the proposed research work is summarized in table 2.

|  |  |
| --- | --- |
| Activity | Deadline |
| Activity | Deadline |
| Literature review | 15.05.22 |
| Database structure design | 31.05.22 |
| Flexible dataflow design for SPA system, that includes the information connected to the pandemic or the other important information for PA | 16.06.22 |
| Research of the databases for predictive analytics on delivery services | 15.06.22 |
| Implementation of the Concept Drift(CD) tweaking methods using C++ | 15.07.22 |
| Recommendation of ML model implementation development | 15.08.22 |
| Developing of the recommendations for the testing and evaluating of the designed system, Testing coordination of DB unit and CD units | 31.09.22 |

Table 2: Schedule of the research work

## 4.3.Data management plan (e.g. Google folders for project logs, reports, literature etc)

The reference literature literature will be stored in local computer.

Logging files and ML logs could be stored in cloud.

## 4.4.Deliverables

The main result of the projects:

- Programming module for CD tweaking;

- Databases links for Predictive Analytics in post COVID-19 times;

- Literature review;

- Recommendation for the creation of SPA;

- Report on the database structure and dataflow for the system;

- Report on the unit and validation tests of the system units.

# 5.Bibliography

[*https://www.cio.com/article/228901/what-is-predictive-analytics-transforming-data-into-future-insights.html*](https://www.cio.com/article/228901/what-is-predictive-analytics-transforming-data-into-future-insights.html)

[*https://www.marketsandmarkets.com/Market-Reports/predictive-analytics-market-1181.html*](https://www.marketsandmarkets.com/Market-Reports/predictive-analytics-market-1181.html)

[*https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ordering-in-the-rapid-evolution-of-food-delivery*](https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ordering-in-the-rapid-evolution-of-food-delivery)

*https://www.iguazio.com/blog/concept-drift-deep-dive-how-to-build-a-drift-aware-ml-system/*

Kadwe, Yamini & Suryawanshi, Vaishali. 2015. A Review on Concept Drift. IOSR Journal of Computer Engineering. 17. 20-26. 10.9790/0661-17122026.

Losing, V., Hammer, B., & Wersing, H. 2018. Incremental on-line learning: A review and comparison of state of the art algorithms. Neurocomputing, 275, 1261-1274.

Žliobaitė, I., Pechenizkiy, M., & Gama, J. 2016. An Overview of Concept Drift Applications. //DOI:[*10.1007/978-3-319-26989-4\_4*](https://doi.org/10.1007/978-3-319-26989-4_4)

Kanoun, Karim & Schaar, Mihaela. (2015). Big-Data Streaming Applications Scheduling with Online Learning and Concept Drift Detection. 1547-1550. 10.7873/DATE.2015.0786.

https://github.com/abifet/adwin

*https://github.com/rsdevigo/pyAdwin*

Bifet, Albert & Read, Jesse & Pfahringer, Bernhard & Holmes, Geoff & Žliobaitė, Indrė. 2013. CD-MOA: Change Detection Framework for Massive Online Analysis. 92-103. 10.1007/978-3-642-41398-8\_9.

Baier, Lucas & Hofmann, Marcel & Kühl, Niklas & Mohr, Marisa & Satzger, Gerhard. 2020. Handling Concept Drifts in Regression Problems - the Error Intersection Approach.

Baier, Lucas & Reimold, Josua & Kühl, Niklas. (2020). Handling Concept Drift for Predictions in Business Process Mining. 76-83. 10.1109/CBI49978.2020.00016.

Zenisek, Jan & Holzinger, Florian & Affenzeller, Michael. (2019). Machine Learning based Concept Drift Detection for Predictive Maintenance. Computers & Industrial Engineering. 137. 106031. 10.1016/j.cie.2019.106031.

Straat, Michiel & Abadi, F. & Kan, Z. & Göpfert, Christina & Hammer, Barbara & Biehl, Michael. (2022). Supervised learning in the presence of concept drift: a modelling framework. Neural Computing and Applications. 34. 1-18. 10.1007/s00521-021-06035-1.

Bahrami Bidoni, Zeynab & Montreuil, Benoit. (2021). Predictive Demand Modeling for New Services in Hyperconnected Urban Parcel Logistics.