DATA SCIENCE PRODUCT DEVELOPMENT

WORD COUNT WITHOUT REFERENCE: 1892

WORD COUNT WITH REFERENCE:2131

Contents

[**INTRODUCTION** 1](#_Toc103410962)

[**Product Design Section** 1](#_Toc103410963)

[**Data source and theme selection and specification** 1](#_Toc103410964)

[**Application domain/end user’s requirements analysis** 2](#_Toc103410965)

[**Product functional and non-functional requirements specifications** 2](#_Toc103410966)

[**Functional requirements** 2](#_Toc103410967)

[**Non-Functional requirements** 2](#_Toc103410968)

[**Product software architecture design** 2](#_Toc103410969)

[**Product use case specifications** 3](#_Toc103410970)

[**Product Development Section** 3](#_Toc103410971)

[**Software and Hardware Tools** 3](#_Toc103410972)

[**Product development software engineering methodology** 4](#_Toc103410973)

[**System testing method** 4](#_Toc103410974)

[**User evaluation plan and methods** 4](#_Toc103410975)

[**Project Management Section** 4](#_Toc103410976)

[**Time management with Gantt Chart** 4](#_Toc103410977)

[**Risk assessment on personal information protection and data security/governance** 6](#_Toc103410978)

[**Quality control on software development** 6](#_Toc103410979)

[**Basic Customer/User relationship management** 6](#_Toc103410980)

[**Basic Product marketing strategy** 7](#_Toc103410981)

[**Conclusion** 7](#_Toc103410983)

[**REFERENCES:** 8](#_Toc103410984)

# **INTRODUCTION**

Product development defined by [V. Krishnan](https://pubsonline.informs.org/action/doSearch?text1=Krishnan%2C+V&field1=Contrib) and [Karl T. Ulrich](https://pubsonline.informs.org/action/doSearch?text1=Ulrich%2C+Karl+T&field1=Contrib) is, “the transformation of a market opportunity and a set of assumptions about product technology into a product available for sale” (Krishnan and Ulrich, 2001). It basically refers to the stages, steps and processes involved in bringing a product from an idea or concept level into the market for sale. This process generally starts with identification of a market need or business problem to be solved and ends with development and release of a Minimum Viable Product (MVP) for the users to test. In this project, following the product development processes, we have designed and developed a product prototype (MVP) called Autism Spectrum Disorder Detector, a product to help doctors better diagnose/detect autistic individuals.

# **Product Design Section**

## **Data source and theme selection and specification**

The data science product developed aims to help doctors and other medical personnel better diagnose autism spectrum disorder. This Disorder is a neurodevelopmental disability characterized by a deficiency in social communication and social interaction, in addition to restricted, repetitive patterns of behavior (centers for disease control and prevention, 2022). The disorder is also quite difficult to diagnose due to the lack of a defined medical test for it (centers for disease control and prevention, 2022), we therefore build this product to help doctors and other personnel in the health sector to better diagnose/detect individuals with autistic spectrum disorder. This will help ensure that people are not wrongly diagnosed with the disorder, while likewise providing an effective and efficient approach to detect those with the disorder and ultimately allowing them to get early health care.

The Product was built using a data science approach, majorly of the machine learning type. It is well known from literature that the performance of a machine learning (ML) model is highly dependent on the quality of the data (Jain, A., et. al., 2020), thus in order to develop an efficient machine learning model that helps predict autism disorder, the first step was to source the appropriate data set.

The final dataset used in developing the model was gotten by combining three datasets; Autism-Adolescent-Data, Autism-Adult-Data and Autism-Child-Data which were all downloaded from the UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php.> The final dataset contains 1100 rows and 22 attributes which includes: ID, Age, Gender, Ethnicity, if born with Jaundice, Family member with PDD, who is completing the test, Country of Residence, Used the screening app before, Screening method type, A1\_score, A2\_score , A3\_score , A4\_score , A5\_score , A6\_score , A7\_score, A8\_score, A9\_score, A10\_score, Result and Class (Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K., & Islam, M. N, 2019). A1\_score, A2\_score, A3\_score, A4\_score, A5\_score, A6\_score, A7\_score, A8\_score, A9\_score and A10\_score is a set of ten questions asked to determine Autism Spectrum Quotient (AQ)/ autistic symptoms (Muthisamy, T., Jayabalan, M. and Rana, M., 2020).

## **Application domain/end user’s requirements analysis**

This software has been designed for use majorly in the diagnosis of autism spectrum disorder, it thus a health-based software. The major end users of this product are doctors and other health/social workers however, the product have also been designed in such a way that allows people who do not fall into this category to be able to use it without difficulty. An end user requirement analysis thus shows that anyone above the age of 15 alongside doctors/health/social workers especially those in care of Autistic Patients will be able to use the product successfully.

## **Product functional and non-functional requirements specifications**

### **Functional requirements**

The product has been developed as a web application with a navigation side bar to the following pages:

* A nice Home/landing page that gives brief introduction on autism spectrum disorder and the purpose of the product.
* An Analytics page that gives insights and fun facts into the dataset. This page also utilizes the use of charts and graphs to communicate these insights and fun facts.
* The Predictor page that takes users inputs based on some questions asked. This page also includes a “Make Prediction” button which when pressed runs a machine learning model in the background and returns a result panel showing if person is autistic or not alongside a pie chart distribution of autism/Not autistic distribution.
* A Questionnaire page is also integrated into the Predictor page. This page shows AQ-10 questions, a set of ten questions asked to determine Autism Spectrum Quotient (AQ)/ autistic symptoms.

While the Home, Analytics and Questionnaire pages are non-interactive, the Predictor page is highly interactive.

### **Non-Functional requirements**

The Non-functional requirements include: Portability, Maintainability, Usability, Availability and Performance

## **Product software architecture design**

The product architecture design employed the modular approach of product design. This approach entails organizing the product into a number of modules, each module performing a specific function that contributes to the overall functionality of the product (Bonvoisin, J., Halstenberg, F., Buchert, T. and Stark, R., 2016). The product was built from the following distinct functional units/components:

* An Analytical and Machine learning model Unit to extract the insights in the dataset and develop a model to predict autism spectrum disorder.
* A web app unit that displays the result of the analysis and uses the model built to make predictions for users.

## **Product use case specifications**

A typical sequence of events that will occur for users of the web system will be to click on the web app link which lands them on the home web-page of the web system where they read up details of autism spectrum disorder, thereafter they can either go to the Analytics page or Predictor page. From the Analytical page, they are able to read up Insights and fun facts about autism spectrum disorder while on the Predictor page the following steps takes place:

* User answers a set of 20 questions, which also includes 10 AQ-10 questions to be answered using the Questionnaire page.
* After Answering the questions, user presses the Button, “Make Prediction” at the bottom left corner of the page.
* The software runs and returns a result for the user to view. The result tells the user if he is either autistic or not. The result also communicates the probability of the user being autistic or not with the aid of a pie chart.

# **Product Development Section**

### **Software and Hardware Tools**

The programming language of choice for the development of the product is the python programming language (Van Rossum, G. & Drake, F.L., 2009) while the mode of deployment for the product is the Web. Python was chosen because it is open sourced and have a vast library support that can handle all the tasks needed to build the product. The libraries/packages used includes: Pandas, Scikit learn, Matplotlib and Seaborn.

The web was chosen as a mode of deployment for ease of accessibility and deployment. Another python package/library called streamlit was used in the development of the web app product.

### **Product development software engineering methodology**

There are different forms of development methodology in software development, each with its own advantages and disadvantages. The development of this data science product utilizes a hybrid approach of agile methodology and waterfall methodology. This hybrid approach entails breaking the overall product development into phases; Data gathering, Analytics, Model development, Web development and continuously iterate through this phase until the right product is achieved.

## **System testing method**

Both functional and non-functional methodology testing forms were adopted. A combination of unit testing, integration testing and systematic testing was done. The respective units that make up the overall system was tested individually to ensure they work as expected before being integrated together. The over-all system formed from the integration of the respective units was likewise tested to ensure the product runs as required, performs all necessary actions and returns expected results. Non-functional tests such as usability, compatibility and performance testing were also carried out.

### **User evaluation plan and methods**

The evaluation of the product will be performed using a beta-testing approach. This would involve getting the users to use the product and getting their feedbacks. The product will also be monitored while it is being used to ascertain all functionality works.

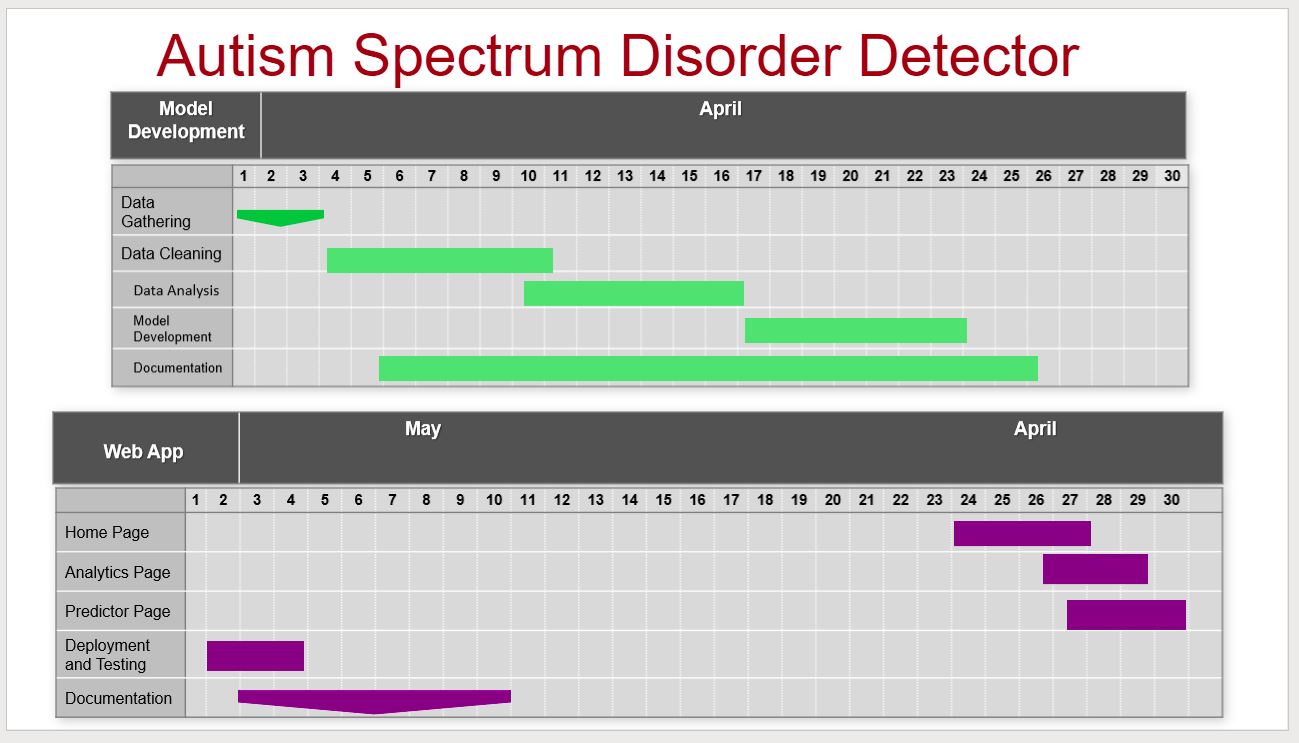
# **Project Management Section**

## **Time management with Gantt Chart**

**Product Development – Schedule of Tasks**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Task** | **Hours** | | **Start-Date** | **End Date** | **Resources** | | **% Complete** |
| Data Gathering | 3 | | 4/1/2022 | 4/3/2022 | UCL repository | | 100 |
| Data Cleaning | 10 | | 4/4/2022 | 4/10/2022 | Python (Pandas) | |  |
| Data Analytics | 10 | | 4/10/2022 | 4/16/2022 | Python (Pandas, Matplotlib/seaborn) | | 100 |
| Model development | 15 | | 4/17/2022 | 4/23/2022 | Python (Scikit-learn) | | 100 |
| Web App Development | 15 | | 4/24/2022 | 4/30/2022 | Python (streamlit) | | 100 |
| Web App Deployment | 5 | | 5/2/2022 | 5/4/2022 | Python (streamlit cloud) | | 100 |
| Documentation | 10 | | 5/5/2022 | 5/10/2022 | MS word | | 100 |
| Testing | | 10 | 5/11/2022 | 5/13/2022 | -------------- | 100 | |

**Gnatt Chart**



## **Risk assessment on personal information protection and data security/governance**

Data security threats remain a prevalent issue in various links of the data industry chain, such as data generation, gathering, processing, and sharing (Sun, Zhang and Fang, 2021). Thus, the need for risk assessment on personal information protection and data security for any data science product being developed. The autism spectrum disorder detector has low risk on exposing users' personal information because the infrastructure of the product does not allow storage of user's data, all users data are erased immediately after usage of the product. This approach efficiently mitigates the risk of data security threats and highly protects users' personal information.

## **Quality control on software development**

Software quality control are procedures performed to plan and control product quality during development ensuring the product meets its quality goals. The quality control of the developed product prototype was conducted by running the web app using various possible set of inputs the user could plug in and seeing the results. The end result of the quality control assessment shows the product greatly conforms to the expected requirements and can be safely deployed for use (altexsoft, 2022).

## **Basic Customer/User relationship management**

Customers involvement in product development is one of the most important resources to building a successful product. it ensures customers need is continually satisfied (Nazari-Shirkouhi, 2015). We plan to implement the Collaborative Customer Relationship Management technique for this product. This technique involves using communication methods (such as email and telephone) between the developer and customers.

## **Basic Product marketing strategy**

### Marketing strategies are plans organizations uses to sell their product and services, they are hence another necessary aspect to have a successful product or service. The market strategy focus for this product is to offer a Pre-Launch Giveaway i.e., identifying the sectors, companies, persons that our product will be of great benefit too and allowing them to use the product for free for few months. Other plans include: Facebook Advertising, Google ads email marketing, customers feedback loop and remarketing.

# **Conclusion**

In this project we have developed a product prototype for health and social workers to help better detect autism spectrum disorder in people. The design process, development process and management process for the prototype was also critically documented. The pipeline for the product prototype development started with data gathering, followed by cleaning, analysis, model development and deployment with a web app. In summary, the project provides an effective and efficient approach to detect autism traits for different age groups since diagnosing autism traits is quite costly and could be a difficult process. A major limitation identified with the prototype however is that it is not designed for people below three years due to lack of training dataset for people in that category.

# **REFERENCES:**

1. Krishnan, V. and Ulrich, K., 2001. Product Development Decisions: A Review of the Literature. Management Science, 47(1), pp.1-21.
2. 2022. [online] Available at: <https://www.cdc.gov/ncbddd/autism/facts.html> [Accessed 10 May 2022].
3. Jain, A., Patel, H., Nagalapatti, L., Gupta, N., Mehta, S., Guttula, S., Munigala, V. (2020). Overview and Importance of Data Quality for Machine Learning Tasks. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.
4. Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K., & Islam, M. N. (2019). A Machine Learning Approach to Predict Autism Spectrum Disorder. 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE).
5. Muthisamy, T., Jayabalan, M. and Rana, M., 2020. Analysis of Autistic Spectrum Disorder Screening Data for Adolescents. International Journal of Current Research and Review, 12(19), pp.23-30.
6. Bonvoisin, J., Halstenberg, F., Buchert, T. and Stark, R., 2016. A systematic literature review on modular product design. Journal of Engineering Design, 27(7), pp.488-514.
7. Van Rossum, G. & Drake, F.L., 2009. Python 3 Reference Manual, Scotts Valley, CA: CreateSpace.
8. Sun, L., Zhang, H. and Fang, C., 2021. Data security governance in the era of big data: status, challenges, and prospects. Data Science and Management, 2, pp.41-44.
9. 2022. [online] Available at: <https://www.cdc.gov/ncbddd/autism/facts.html> [Accessed 10 May 2022].
10. Nazari-Shirkouhi, S., 2015. Investigating the effects of customer relationship management and supplier relationship management on new product development. Tehnicki vjesnik - Technical Gazette, 22(1), pp.191-200.