Store Layout and Shelf Planning

An Artificial Intelligence Based Approach

1st Eduardo Silva
Data Engineer
Critical TechWorks
Porto, Portugal
1180881@isep.ipp.pt

2nd Michel Silva
Engineering Lead
Farfetch
Porto, Portugal
1200169@isep.ipp.pt

3rd Miguel Costa
Software Engineer
Critical TechWorks
Porto, Portugal
1210064@isep.ipp.pt

4th Rui Oliveira
RPA Developer
Sonae MC
Porto, Portugal
1210067@isep.ipp.pt

***Abstract*—With client satisfaction becoming more of a concern for retail companies as time goes by, the need for solutions and strategies that help provide clients with the best shopping experience as never been higher.**

In a physical store, few things are as important to clients as the store's layout and product placement, with a familiar yet efficient layout being in their best interest, while a strategic placement of products and categories is performed by merchandising teams, helping boost the company's sales and revenue.

In the following article, the team proposes a solution that seeks to combine the modular and proven effectiveness of genetic algorithms on store planning problems with the knowledge extracted by machine learning techniques, such as the apriori algorithm and neural networks, from historical sales data, enabling the automation of the store planning process, coming up with plans tailored to the clients' habits and expectations.

Despite currently having some limitations, such as the setup complexity and poor reaction to sudden changes to the business context (e.g. a new pandemic), the integration with machine learning techniques, historical data orientation, modularity and flexibility displayed by the solution is promising and a signal that this innovative approach is worth pursuing and be built upon, in order to reach its full potential and be of the utmost value for retail companies all around the world.

***Keywords*—Retail, Sales, Store, Layout, Planning, Artificial Intelligence, Machine Learning, Regression, Apriori, Genetic, Neural Networks.**

I. Introduction

The key indicator of success for retail companies all around the world is clear: **sales and profit**, with customer relationship and satisfaction emerging as important factors to manage in order to come out on top [1].

As such, many are the strategies and techniques being adopted by retail companies to monitor and react to customer activity and feedback, ranging from customer relationship management (CRM) software to artificial intelligence (AI) based solutions [2] [3].

In a previous project, the team had the opportunity to work on such a solution: a customer satisfaction diagnosis tool, supported by an expert system that analysed client activity in an e-commerce platform, generating action plans to recover the confidence of less satisfied clients in the company's services.

This time around, the team shifted its attention from e-commerce to physical stores, having identified store layout and shelf planning as a relevant and current problem for retail companies all around the world, which is currently handled mostly manually by merchandising departments, taking up a significant amount of time and effort for them to handle on a frequent basis.

For this reason, the team proposes a solution that is able to automatically plan and evaluate a physical store layout based on historical data (e.g. last 5 years) and a set of heuristics that reflect the company's merchandising philosophy (e.g. minimize distance between products that are frequently bought together).

In turn, this allows merchandising departments to spend more time on other important tasks at hand, while clients are strategically directed to high-priority products, driven to impulse purchases and provided with a sense of empowerment that, in turn, enhances their retail experience [1].

Given the pandemic context that this project was developed in, the team was able to obtain a sales dataset from January of 2019 to May of 2021, covering both normal circumstances and the COVID-19 pandemic.

As such, the team took the opportunity to analyse and develop a solution around the change in behaviour and habits displayed by clients during a pandemic when compared to a more typical set of circumstances.

In the next few chapters, the reader will be presented with a state of the art analysis of the problems and techniques applied in the proposed solution (i.e. store planning, product/category association and sales forecasting) and how each of them contribute to the problem resolution.

Afterwards, an overview of the proposed solution will be presented, along with its benefits and disadvantages when compared to other existing solutions.

Lastly, the various phases of the solution's development will be followed, loosely based the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, all the way from the business and data understanding to the implementation, evaluation and deployment phases, followed by some closing remarks by the team after this project's conclusion.

II. State of the Art

In order to produce a state of the art solution that is able to not only automatically come up with a plan for a given physical store, but also consider client habits and product demand based on historical sales data, the application of state of the art techniques also becomes a requirement.

As such, in the following sub chapters, the reader will be presented with a brief analysis of (i) store planning, (ii) product and category association, (iii) sales forecasting, how each of these subjects integrate with each other and what techniques are being applied by researchers and retail companies in the present day to tackle the problems involved in each of them.

A. Store Planning

Store planning, the central focus of this project, emerges as topic of significant importance for retail companies with physical stores and, to a certain extent, e-commerce platforms as well, although the planning of the latter is based on different techniques, such as UX design and recommendation systems.

As mentioned before, physical store planning is usually performed by merchandising departments, through careful analysis of customer habits and behaviour, historical sales data and, to a certain extent, human psychology [1].

With the right dataset and artificial intelligence techniques, it is possible to implement a solution that covers these subjects, with the latter being naturally more difficult than the others, given the uncertainty surrounding human psychology and nature.

Nowadays, genetic algorithms emerge as the most commonly applied technique to achieve automatic store planning, given its characteristics and overall effectiveness.

Being originally proposed by John H. Holland in 1975, in the book titled "Adaptation in natural and artificial systems" [4], genetic algorithms take advantage of the knowledge around evolution and its concepts, such as survival, fitness, reproduction and generations, in order to find an optimal solution within a large set of possible values and combinations, while also providing implicit parallelism enabled by the evolutionary analogy [5].

Store layout and shelf planning perfectly fits this description of a problem with a large set of possible solutions, whose value may vary according to a predefined set of heuristics and rules that are in the best interest of retail companies.

As such, although it has been commonly used to solve other problems in retail for many years such as staff scheduling [6] and credit risk assessment [7], genetic algorithms have recently emerged as an effective technique to execute physical store planning, such as product and space allocation in shelves [8] [9] [10], coming up with solutions that offer the most value for the company (i.e. maximizes sales and revenue), while also achieving it within an acceptable time frame [8].

B. Product and Category Association

When browsing platforms such as Amazon or E-Bay, it is common to find sections titled "People also bought X" or "Products frequently bought together", generated by the platforms' product recommendation system, usually supported by an apriori algorithm.

First introduced in 1994, by Rakesh Agrawal and Ramakrishnan Srikant, in the scientific paper "Fast Algorithms for Mining Association Rules", the apriori algorithm emerged as significantly more efficient alternative to the techniques being applied at the time to extract item association rules from transactional datasets, outperforming them by factors ranging from three for small problems to more than an order of magnitude for large problems [11].

Given the relevancy of product association for the business understanding [12] [13] and the definition of product selling strategies [14], the apriori algorithm is frequently applied by retail companies in their technical reports and online platforms.

By integrating the business and client understanding, obtained through the apriori algorithm, with the genetic algorithm, it is possible to mirror a product recommendation system in a physical level and satisfy heuristics such as the distance minimization between products frequently bought together (e.g. clients expect to find fruit in close proximity to vegetables), which ultimately helps towards the maximization of sales and revenue of the company.

C. Sales Forecasting

The ability to accurately predict future product sales and the impact that business decisions have on them has been a dream of retail companies for as long as retail exists.

In recent times, advancements in the field of machine learning and the significant amount of collected data have allowed retail companies to implement their own sales forecasting solutions, resorting to regression algorithms such as polynomial regression and neural networks, with varying levels of success depending on the quality and quantity of the data collected by the company.

The first of the aforementioned techniques, polynomial regression, despite being one of the oldest machine learning algorithms, originally introduced by Sir Francis Galton in 1885 while examining the relationship between fathers' and sons' heights, is still among the most common statistical techniques for investigating and modeling relationships among variables [15], including sales forecasting [16] [17].

As an alternative, neural networks have recently emerged as a powerful non-linear regression algorithm [18] [19] [20], with its ability to identify patterns being particularly suited to the uncertain and irregular nature of a sales forecasting problem, despite also being commonly applied to more complex problems in retail such as product recognition [21] and e-commerce chat bots [22].

While analysing the application of regression algorithms to sales forecasting, researchers have found that variables such as season, culture and region may amplify the complexity of the problem, but at the same time they may also allow the models to more easily find patterns in the product sales throughout a year [16] [18].

On the other hand, researchers have also found that seasonal variables may introduce a certain bias into the predictions, which may not allow the model to generalize as well to unknown circumstances, as it will follow the seasonal trends [18]. As such, deseasonalization techniques may be applied if further improvement and generalization is required.

By integrating sales forecasting with the genetic algorithm, it becomes possible to formulate a product and category demand ranking, allowing the genetic algorithm to better evaluate a layout's aptitude to satisfy client demand in a specific time of year (e.g. around Christmas, clients are expecting to find toys, chocolates and Christmas decoration in store shelves).

III. Proposal

In the following chapter, a brief overview of the solution and its main characteristics will be presented, complemented with an analysis of its benefits and limitations when compared to other existing solutions.

A. Overview

Based on the state of the art techniques identified in the previous chapter, the team proposes a two-phased solution with two similar genetic algorithms at its core, but with distinct roles: the first one is responsible for the layout/section planning, by associating product categories to each section of the store, followed by a second one responsible for planning the product placement in the shelves.

When executing the solution, the user is asked to provide a file containing information about the store layout (i.e. section placement, number of shelves and characteristics of each of them, etc.), a date indicating the time of the year when the plan is desired and an indication about the pandemic context around that time.

How each of these inputs impact the plans formulated by the solution will be explained with further detail in chapter 4.

An execution of this solution essentially comprises of the following steps:

1) The user provides a file with information about the store layout, the desired date for the plan and if there is a pandemic going on at that time;
2) The first genetic algorithm is executed, generating a plan for the store layout/sections, by associating a product category to each section of the store;
3) The second genetic algorithm is then executed, for each section of the store, generating a plan for the product placement in each section's shelves;
4) The resulting store layout and shelf plans are then provided to the user.

It's worth noting that both of the genetic algorithms are supported by the association rules obtained through the apriori algorithm and a product/category sales ranking, generated by the sales forecasting models.

This integration allows both algorithms to better evaluate a plan's value for the store and its clients, by having heuristics such as the distance minimization between associated products/categories and the presence of the most sought products.

In fact, other heuristics also end up being positively impacted by this integration, since both the association rules' confidence level and min-max normalization of the sales ranking result in a set of values between 0 and 1, that are representative of the relative relevancy of a given association rule or product/category for the problem at hand, which can then be used to give a relative priority to certain items within an heuristic's evaluation (e.g. when evaluating a shelf plan, it may be desired to have the most relevant products at eye level).

To summarize this topic, the diagram at fig. 1 displays an overview of how each of the aforementioned components are integrated with one another, with both genetic algorithms having a similar integration with the other components, despite the first one being more focused on product categories and the latter in the products themselves.

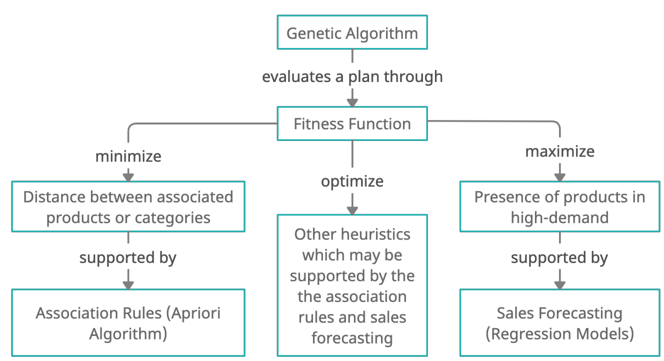

Fig. 1. Proposed Solution - Component Integration

B. Benefits

When compared to other existing solutions, the solution proposed in this article contains a set of benefits that help it stand out on its own as an innovative approach to automate store layout and shelf planning.

On the one hand, the modular nature of the genetic algorithms allows retail companies to implement their own set of customized heuristics, so they better reflect the company's merchandising philosophy, resulting in plans better suited for their clients.

On the other hand, the integration of machine learning techniques with the genetic algorithms endows the solution with knowledge obtained directly from historical data, which, in turn, results in plans tailored for the clients' habits and expectations.

Another aspect where this solution excels is its flexibility in the supported store layouts, being able to dynamically load a store layout/schematics at execution time, directly from a JSON file that contains information such as section placement, number of shelves in each section, among other data.

Lastly, it must be noted that the solution is able to adapt to different seasons/times of the year and even to a pandemic context, with the latter still being particularly relevant in 2022, as client behaviour and habits change over time and according to the context they live in.

C. Limitations

With the aforementioned benefits, there also come a few limitations when compared to other existing solutions.

On the one hand, the reliance on historical sales data can prove to be difficult for smaller retail companies to manage. The setup process and the amount of pre-processing required in order to guarantee the quality of the data can also be difficult, even for bigger supermarket chains.

On the other hand, this solution doesn't react well to sudden changes in circumstances (e.g. a new pandemic or a war), requiring more data about the context surrounding the business in order to achieve that.

IV. Development

In the following chapter, a summary of the various phases of development undertaken during this project's duration will be presented.

It's worth noting that the chapter's structure is loosely based on the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, beginning with a business and data analysis, proceeding with the data preparation and solution implementation, and finishing with its evaluation and deployment.

A. Business Understanding

In retail companies, merchandising departments are usually assigned to come up with a plan for the physical stores, always aiming to strike a balance between client satisfaction, by keeping the layouts familiar and efficient, and the company's best interests, such as strategic product and section placement, that aim to drive clients to the highest-priority products and to impulse purchases, ultimately generating more revenue.

This process is usually split into different phases, with the team opting to cover two of them in this project: (i) layout/section planning and (ii) product placement in the shelves, preferably executed in this order.

In the first phase, the main concern of the plan is to attribute product categories to each section of the store, while keeping in mind categories that are in high-demand, seasonal or even usually associated/bought together.

A classic example of impulse purchasing, as a result of placing two commonly associated categories close together, are fathers that go to the supermarket to buy diapers for their baby, usually tending to buy a pack of beers on the way to checkout.

In the second phase, the main concern shifts to the product placement on the shelves, having already attributed a product category to each section.

During this phase, a set of heuristics have to be kept in mind, such as the placement of the most relevant products at eye level and, depending on the store's merchandising philosophy, in the middle or at the edges of the section, in case they want clients to check all the products before reaching the one they want or they want to catch their attention even before they enter said section.

B. Data Understanding and Preparation

For this project, the team was able to retrieve a dataset with sales from January of 2019 to May of 2021, relative to a south american supermarket with around 2.5 million annual sales, totalling around 6GB of data and about 40 million entries.

A first look at the data reveals the many attributes present in the dataset, such as receipt ID, quantity and value of the purchase, four levels of product description, among others, with only the following being deemed relevant by the team for the problem at hand:

- **Ticket ID -** Receipt/Basket identifier;
- **Date -** Date of the purchase;
- **Value -** Total cost of the product in the receipt;
- **Quantity -** Quantity/Units of the product that were bought;
- **Category -** Category of the purchased product (e.g. Vegetables);
- **Product -** Product designation (e.g. Tomato).

Having filtered the relevant attributes, further pre-processing had to be applied on the data, starting with the treatment of missing data. Luckily, the data missing from the dataset was insignificant and considering the amount of entries in the dataset, it was decided that entries with missing data would simply be dropped.

Afterwards, as this dataset refers to a south american supermarket, an adaptation of certain aspects such as language and currency was required, from spanish and uruguayan pesos to english and euros, respectively.

For the language adaptation, a manual translation of 53 category and 578 product labels had to be performed by the team, while the conversion from uruguayan pesos to euros only required a multiplication of the sales' values by the conversion rate of uruguayan pesos to euros (around 0.020 by the time this article was written).

Regarding outliers, it was decided that categories and products whose sales were not significant should be dropped, with the team settling on the median as a reference value, in order to attain a more generalized but still rich and representative dataset, while also potentially improving the quality of machine learning models that are trained with this dataset.

Finally, to further reduce the amount of memory taken by the dataset and facilitate queries, the team decided to separate the dates into separate day, month and year attributes, dropping the hours at which a sales was made, as it was not really necessary for the solution implementation.

In table I, a summary of the data pre-processing is presented, demonstrating that the complexity of the problem has been significantly reduced, by dropping insignificant categories and products, while still retaining the majority of the dataset entries, resulting in a 66% reduction in file size, which should improve the performance and speed at which the team can experiment with the data.

TABLE I
DATA PRE-PROCESSING RESULTS

	Before	After	Reduction
No. Categories	129	53	60%
No. Products	2 069	578	72%
No. Receipts	5 141 676	4 844 487	6%
No. entries	42 244 350	37 243 405	12%
Dataset size	Around 6GB	Around 2GB	66%

Having cleaned the data, the team proceeded with an exploratory data analysis, where a few interesting insights were found, namely the business impact caused by the COVID-19 pandemic, around March of 2020.

In fig. 2, a line plot is shown, displaying the evolution of the supermarket monthly sales from January of 2019 to May of 2021, with the impact of the COVID-19 pandemic being quite evident, by the sudden drop in sales by the start of 2020, reaching a global minimum in April of 2020.

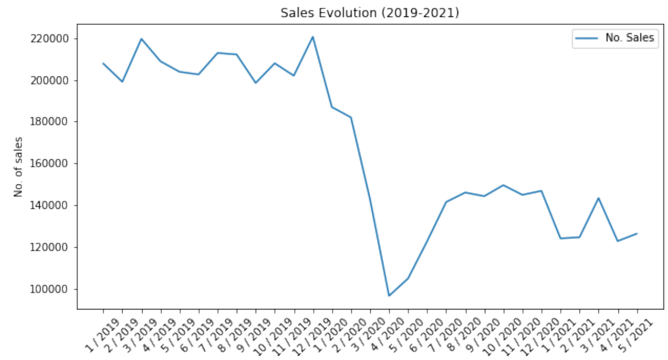

Fig. 2. Sales evolution from January of 2019 to May of 2021

From the plot it is also possible to observe how the monthly sales have not recovered ever since the start of the pandemic, recovering to around half of the monthly sales observed in 2019, as shown in table II:

TABLE II
MONTHLY SALES EVOLUTION

Year	No. of Sales
2019	207968.67
2020	142349.42
2021	128134.00

By decreasing the granularity and comparing some of the category sales recorded by the company in the same time frame, a few interesting insights turn up in the line plot shown in fig. 3.

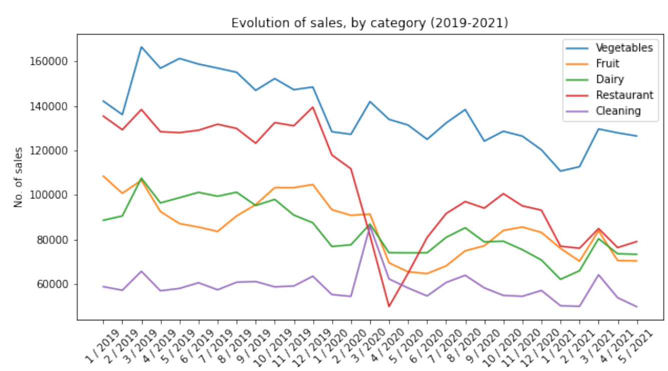

Fig. 3. Sales comparison between categories

From the previous graph, it is interesting to check that the restaurant category took a big hit at the start of the pandemic, taking a few months to recover, but never coming close to the vegetables as it once did in 2019.

In the other hand, hygiene products had a significant boost in sales at the start of the pandemic, most likely due to the panic caused by COVID-19. It is worth noting that during a new wave, at the start of 2021, another spike in hygiene sales is visible, despite being smaller.

Finally, products such as vegetables, fruit and dairy, being essential categories for the day-to-day life, didn't take as much of a hit, despite also having their sales slowly decrease over time.

Further analysis into the data also showed that client habits do change throughout the year, as expected (e.g. around Easter and Christmas, the sales and revenue generated by chocolates and cakes almost doubles).

The analysis performed by the team shows that there is potential for the solution to tailor its plans according to client habits and expectations throughout a year, depending on factors such as season and the proximity of holiday.

Furthermore, given the change in behaviour demonstrated by clients during the pandemic and relevancy of store planning in the pandemic context that the world still faces in 2022, it would be a missed opportunity not to implement a solution that is capable of adjusting its plans according to the context around the business and its clients.

C. Implementation and Evaluation

In the following sub chapter, a brief overview of the implementation and evaluation process for each component of the proposed solution will be presented.

- Apriori algorithm

As previously mentioned, by applying the apriori algorithm to the sales dataset, it is possible to identify products and categories that are commonly bought together, helping the genetic algorithm value plans that have these items close to each other.

Out of the three components, the apriori algorithm was the easiest to setup and apply, as it only requires the creation of a pivot table with the receipt ID's and categories/products as the table entries, having the corresponding sales count as the entry values, encoded as 1 or 0, in case a receipt contains that item or not.

Given the computational complexity of the apriori algorithm, particularly in terms of memory, the team had to find a strategy to handle the significant amount of data present in the dataset, otherwise the algorithm kept running out of memory during its execution.

As it was also interesting to analyse if clients had different behaviours in different times of the year, the team decided to apply the algorithm for each quarter of a year, which significantly reduces the necessary memory to run a single instance of the algorithm, as the number of receipts is smaller, but also allows the genetic algorithm to consider different association rules depending on the time of year for which the plan was asked for.

In tables III and IV, some of the most interesting association rules found by the team are presented, with the first table showing rules identified in the first quarter of 2019 and the second one in the first quarter of 2020, making it possible to analyse how people's habits changed with the start of the pandemic.

TABLE III
EXCERPT OF THE OBTAINED ASSOCIATION RULES
(CATEGORIES - Q1 2019)

Antecedents	**Consequents**	**Confidence**
Fruit	Vegetables	63%
Cheese	Dairy	60%
Dairy	Fruit	59%
Meat	Vegetables	52%
Cheese	Bakery	50%

TABLE IV
EXCERPT OF THE OBTAINED ASSOCIATION RULES
(CATEGORIES - Q1 2020)

Antecedents	**Consequents**	**Confidence**
Fruit	Vegetables	65%
Cheese	Dairy	62%
Canning	Vegetables	56%
Hygiene	Dairy	56%
Meat	Vegetables	54%

Right away, it is interesting to verify that in 2020, with the start of the pandemic, canning and hygiene products started being bought together with other everyday products such as vegetables and dairy more frequently, suggesting that people got scared and started to buy more of these products in order to protect themselves.

- Sales forecasting

Having identified both the polynomial regression algorithm and neural networks as effective machine learning techniques for sales forecasting, the team decided to experiment with both techniques and find out which one produced better results for the available data.

Before any experimentation was done, additional data preparation had to be performed in order to proceed with the model training.

New datasets had to be prepared for each category and product in the company's catalog, with the weekly sales throughout the year and new categorical features obtained through feature engineering: seasons and holidays, as they may help the models to more easily identify seasonal patterns (e.g. clients are looking for cold drinks and ice creams during the summer).

By the end of this process, the new datasets present the following structure:

- **Week no. -** Week number in the year (i.e. from 1 to 52);
- **Season -** Season of that week (e.g. in the 1st week of the year, it is Winter);
- **Holiday -** Closest holiday to that week (e.g. in the penultimate week of the year, the 51st, it's Christmas);
- **No. of purchases -** Number of purchases in that week;

It must be noted that the new categorical categories had to be one-hot encoded in order to be used in the regression models' training.

Having completed the data preparation process, each dataset went through a 70%/30% split, in order to obtain the training and test sets. The training process was then executed, resorting to a grid search with a 5 fold cross validation as a means to optimize the model's hyperparameters.

After some experimentation and analysis, the team found neural networks to generally be more accurate than an equivalent polynomial regression model, as shown by the line plot in fig. 4, where the neural network came out on top, with a correlation coefficient of around 0.70, compared to the 0.61 of the polynomial regression model, when predicting the weekly sales of vegetables throughout the year.

Fig. 4. Sales forecasting model comparison

It's worth noting that for this problem in particular, a correlation coefficient close to 1 might not be desired, since the sales count variation is very volatile and, at times, unpredictable. A correlation coefficient close to 1 would mean that the model might be overfitting the data and following every single sudden variation that occurred in the recorded sales.

In this case, the team found that a correlation coefficient between 0.5 and 0.7 might be ideal, as the experiments demonstrated that models with a correlation coefficient in this range tend to be more generalized, ignoring most of the sudden changes in favor of more gradual variations, retaining a better overall representation of reality and adapting better to unknown scenarios.

Having identified neural networks as the more effective technique of the two, the team proceeded with the execution of a training/evaluation pipeline for each product and category in the supermarket's catalog, obtaining a sales forecasting model for each of them, and also for both 2019 and 2020, in order to help the genetic algorithm identify the most wanted products in both normal and pandemic circumstances.

Given the computational complexity of the grid search algorithm, it was decided that a smaller search space had to be considered, as it would take too much time to train 53 neural networks for the categories and 578 for the products.

As a result, the aggregated performance metrics displayed in table V are not as good as they could possibly be, with the correlation coefficient falling short of the desired interval.

TABLE V
EXCERPT OF THE OBTAINED ASSOCIATION RULES
(CATEGORIES - Q1 2020)

	2019 Cat.	2020 - Cat.	2019 - Prod.	2020 - Prod.
Mean MAE	335.08	415.60	61.85	58.48
Mean RMSE	532.90	913.65	107.91	116.19
Mean R^2	0.62	0.42	0.46	0.42

Nonetheless, as a proof of concept, the team believes that these models serve their purpose and help the genetic algorithm identify the most-wanted products at any given time of the year and either in a normal or pandemic context.

- Genetic algorithms

Finally, the genetic algorithms were implemented, with all the association rules and sales forecasting models already available to be integrated in the fitness functions.

It must be noted that a set of auxiliary functions had to be implemented before the implementation of the fitness functions and the heuristics themselves, with the most important being coordinates functions, that help find the center point and calculate distances between the different sections in the store.

Moving on to heuristics, the team wanted to keep this proof of concept as generalized as possible, without any specific store philosophies in mind. As such, the following heuristics were settled by the team as the most relevant and general ones, in which the majority of retail companies would agree:

- Layout/Section planning
  - Hard constraints
    - ∗ Maximize the presence of the most wanted categories;
    - ∗ Minimize the distance between associated categories;
    - ∗ Frozen categories must be in a freezer;
    - ∗ Fresh categories must be in a fridge;
    - ∗ Categories identified for waist-level stands have to associated with the respective sections (e.g. fruits and vegetables);

- Categories identified to be close to the walls have to be associated with the respective sections (e.g. butchery and bakery).
- Soft constraints
  - Minimize the distance between similar categories (e.g. fruit and vegetables);
  - Minimize the distance from seasonal products to the entrance;
  - Categories with most products in the catalog should be placed in sections with more room.

- Shelf planning
  - Hard constraints
    - Maximize the presence of the most wanted products;
    - Minimize the distance between associated products;
    - Most relevant products must be at eye level.
  - Soft constraints
    - Most relevant products should be by the middle of the corridor;
    - Duplicate products should be close to each other.

Regarding the genetic algorithms' parameters, the team found that the layout/section planning usually requires around 2000 generations, while the shelf planning only requires around 300 generations for each section. Despite this, it was decided that in case any of the algorithm instances requires a few more generations to converge to an optimal solution, a margin of 1000 and 200 generations would be granted to each, with a stoppage criteria set for a 200 generation stagnation.

To summarize, the most relevant genetic algorithm parameters are displayed in table VI.

TABLE VI
GENETIC ALGORITHM PARAMETERS

Parameter	**Value**
No. generations (section planning)	3000
No. generations (shelf planning)	500
No. individuals per generation	50
Parent selection method	Steady state selection
Crossover method	Two point crossover
Crossover probability	80%
Mutation probability	5%
Stoppage criteria	200 generation fitness stagnation

With these heuristics and parameters in place, the genetic algorithms were finally ready to execute store layout and shelf planning. In fig. 5 and 6, an example of a store layout and shelf plan are displayed, respectively.

Fig. 5. Example of a store layout/section plan

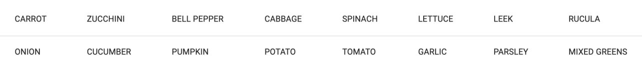

Fig. 6. Example of a shelf plan, for the vegetables section

By analysing the plans displayed in the figures, the team would would say that the majority of sections seem to be in familiar or expected places, although it is difficult to evaluate, as the algorithm may have found insights that are imperceptible to users and lead to the suggested plan, despite some of the decisions seeming odd.

On the other hand, being a multi objective decision making problem, it is possible that the heuristics may be too generalized and the problem requires a more complete set of heuristics in order to perform at its best, being something to look into in future work and retail companies that are looking capture their merchandising philosophy in the genetic algorithms.

D. Deployment

Being a proof of concept, the team found the idea of developing a web application interesting, in order to demonstrate a real use-case of the proposed solution.

As such, the team designed the web application represented by a component diagram in fig. 7, which comprises of a single page application in the front-end and a REST API in the back-end.

The latter allows the front-end to communicate with the genetic algorithms and request a store layout and shelf plan for the parameters specified by the user, with the genetic algorithms, in turn, connecting with the trained sales forecasting models.

It's worth noting that the genetic algorithms and machine learning models were implemented with the PyGAD and scikit-learn libraries, respectively.

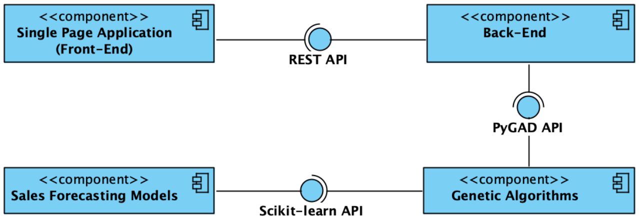

Fig. 7. Web application - Component diagram

In fig. 8, a screenshot of the single page application is presented, showcasing the form that users have to fill in order to execute the planning process, as well as the store layout render, which was made possible through a custom built store rendering module.

Fig. 8. Web application - Single page application

It must be noted that after the planning is finished, the store render in the single page application is then updated with the respective section attributions, as shown in fig. 5.

V. Conclusion

With client satisfaction becoming an increasingly more important factor for retail companies to consider as time goes by, the adoption of solutions and strategies that seek to manage this is also becoming increasingly more crucial for retail companies, in order to provide clients with the best possible shopping experience.

Data-driven, flexible, modular and integrated with machine learning techniques. The innovative approach followed in this project has the potential to be expanded and improved even further, bringing it closer to a solution that every retail company around the world would want deployed in their merchandising departments.

One such improvement would be the support for already existing store plans, with the solution just considering specific sections that the user has identified. This might be useful for stores that are already well established and clients are familiar with, in which case a new plan from the ground up may cause some confusion and an overall bad shopping experience.

By only considering specific sections, the solution could then focus on the placement of seasonal products, such as Easter eggs and Christmas decoration, around the respective times of the year. At the moment, despite the solution already considering seasonal products, there's no specific restriction of where to place those products, only a heuristic that values the proximity to the store's entrance.

On the other hand, in order to mitigate the complexity of the setup and data preparation necessary to deploy this solution, it would be interesting to implement a more generalized and user-friendly service, providing an interface where users are only asked to insert information about the available data (e.g. data types, currency) and the service would then take care of the setup, all the way from the data pre-processing and analysis to the application of machine learning techniques and setup of the genetic algorithms.

Furthermore, given the modular nature of the solution, it would also be in the interest of retail companies to be able to configure their own customized heuristics and rules in the genetic algorithms, with the service providing an easy method to do just that.

All in all, this project might be over, but store layout and shelf planning is a problem that will continue to be studied for many years to come, with new and innovative solutions being proposed all the time, pushing the boundaries of what is believed to be possible in the retail domain.

Hopefully, the solution proposed in this article has contributed to that push, with the team encouraging the scientific community to further investigate and build upon where this solution was left off, allowing it to reach its full potential and be of the utmost value to retail companies all over the world.

References

[1] A. Mishra and R. Vishvas, "Retail shopper empowerment: A consumer-centric measure for store performance," *IIMB Management Review*, vol. 31, no. 1, pp. 20–36, mar 2019.
[2] P. C. Verhoef, "Understanding the Effect of Customer Relationship Management Efforts on Customer Retention and Customer Share Development," vol. 67, no. 4, pp. 30–45, oct 2018.
[3] M. Roberts-Lombard, "Customer retention through customer relationship management: The exploration of two-way communication and conflict handling," *African Journal of Business Management*, vol. 5, no. 9, pp. 3487–3496, may 2011.
[4] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, 1975.
[5] D. E. Goldberg and J. H. Holland, "Genetic Algorithms and Machine Learning," *Machine Learning*, vol. 3, no. 2, pp. 95–99, 1988.
[6] S. Zolfaghari, V. Quan, A. El-Bouri, and M. Khashayardoust, "Application of a genetic algorithm to staff scheduling in retail sector," *International Journal of Industrial and Systems Engineering*, vol. 5, no. 1, pp. 20–47, 2010.

[7] S. Oreski, D. Oreski, and G. Oreski, “Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment,” *Expert Systems with Applications*, vol. 39, no. 16, pp. 12 605–12 617, nov 2012.
[8] K. Czerniachowska, “A genetic algorithm for the retail shelf space allocation problem with virtual segments,” *OPSEARCH*, pp. 1–49, aug 2021.
[9] V. Holý, O. Sokol, and M. Černý, “Clustering retail products based on customer behaviour,” *Applied Soft Computing*, vol. 60, pp. 752–762, nov 2017.
[10] M. Castelli and L. Vanneschi, “Genetic algorithm with variable neighborhood search for the optimal allocation of goods in shop shelves,” *Operations Research Letters*, vol. 42, no. 5, pp. 355–360, jul 2014.
[11] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, 1994.
[12] P. R. Gaikwad, S. D. Kamble, N. V. Thakur, and A. S. Patharkar, “Evaluation of Apriori Algorithm on Retail Market Transactional Database to get Frequent Itemsets,” *Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*, vol. 10, pp. 187–192, jun 2017.
[13] K. Venkatachari and I. D. Chandrasekaran, “Market basket analysis using fp growth and apriori algorithm: a case study of mumbai retail store,” *BVIMSR's Journal of Management Research*, vol. 8, no. 1, p. 56, 2016.
[14] W. B. Zulfikar, A. Wahana, W. Uriawan, and N. Lukman, “Implementation of association rules with apriori algorithm for increasing the quality of promotion,” *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*, sep 2016.
[15] C. A. Ethington, S. L. Thomas, and G. R. Pike, “Back to the Basics: Regression as It Should Be,” pp. 263–293, 2002.
[16] Y. Ni and F. Fan, “A two-stage dynamic sales forecasting model for the fashion retail,” *Expert Systems with Applications*, vol. 38, no. 3, pp. 1529–1536, mar 2011.
[17] A. Krishna, V. Akhilesh, A. Aich, and C. Hegde, “Sales-forecasting of Retail Stores using Machine Learning Techniques,” *Proceedings 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions, CSITSS 2018*, pp. 160–166, dec 2018.
[18] G. C. Aye, M. Balcilar, R. Gupta, and A. Majumdar, “Forecasting aggregate retail sales: The case of South Africa,” *International Journal of Production Economics*, vol. 160, pp. 66–79, feb 2015.
[19] Y. Kaneko and K. Yada, “A Deep Learning Approach for the Prediction of Retail Store Sales,” *IEEE International Conference on Data Mining Workshops, ICDMW*, vol. 0, pp. 531–537, jul 2016.
[20] A. L. Loureiro, V. L. Miguéis, and L. F. da Silva, “Exploring the use of deep neural networks for sales forecasting in fashion retail,” *Decision Support Systems*, vol. 114, pp. 81–93, oct 2018.
[21] M. Hadi Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, “Where to Buy It: Matching Street Clothing Photos in Online Shops,” pp. 3343–3351, 2015.
[22] L. Cui, S. Huang, F. Wei, C. Tan, C. Duan, and M. Zhou, “SuperAgent: A Customer Service Chatbot for E-commerce Websites,” pp. 97–102, 2017.