

AI Product/Service Prototyping for Small Businesses
Feynn Labs
Task-0

Spend Classification Service

Manjunath Kannavalli

12/03/2023

Abstract

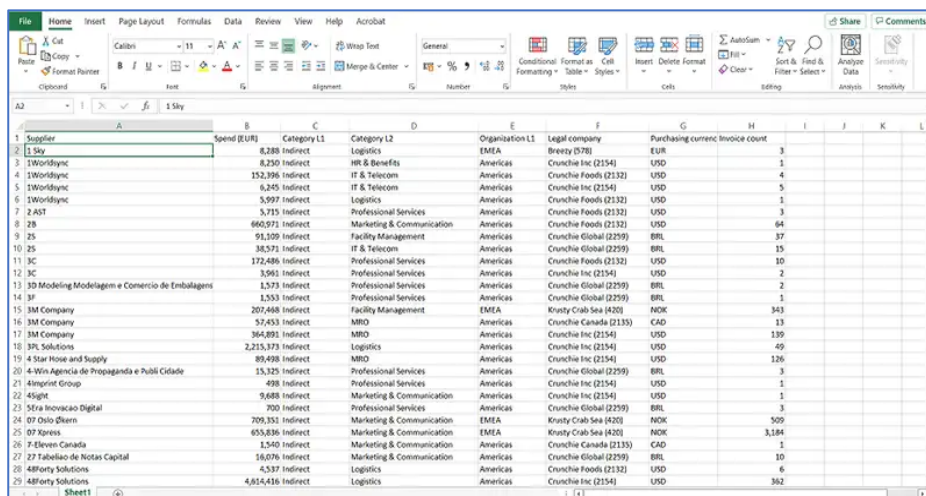
Spend classification is a crucial task for procurement professionals to gain visibility into their spend patterns, identify savings opportunities, optimize sourcing strategies, and monitor compliance. However, manual spend classification is time-consuming, error-prone, and inconsistent. In this project, we propose a machine learning approach to automate and improve the accuracy of spend classification. We use various techniques such as data cleaning, feature engineering, model selection, training, testing, and evaluation to classify spend data into predefined categories based on various attributes such as supplier, product, service, department, etc. We compare our results with other methods or benchmarks and discuss the implications for procurement practice. We also address the limitations of our approach and suggest possible improvements or future work.

1.0 Problem Statement

The problem statement for spend classification in small businesses is to develop a machine learning model that can accurately categorize the various types of expenses incurred by a small business. Small businesses generate a lot of financial data, which can be challenging to manage manually. The current practices of using Excel or BI tool is error prone and time consuming due to human interaction. Excel does not have advanced machine learning capabilities and may require manual coding and formulas to perform complex data analysis and classification.

2.0 Market/Customer/Business Need Assessment

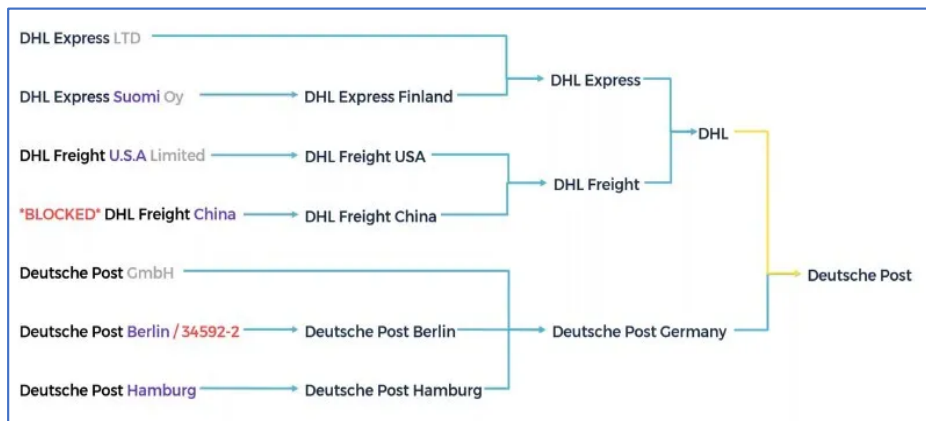
Spend classification is a critical process for businesses to manage their spending and optimize their financial performance. While tools like Excel or business intelligence (BI) software can be used for spend classification, they have some limitations. For example, they require significant manual effort and are prone to human error, which can lead to inaccuracies in spend classification. According to a study by McKinsey, traditional methods of spend classification achieve an accuracy rate of 60% to 70%. In contrast, machine learning-based spend classification algorithms can achieve an accuracy rate of 90% or higher. Additionally, machine learning can speed up the spend classification process, reduce manual effort, and provide deeper insights into spending patterns. Machine learning can also handle large volumes of data and scale to meet the needs of growing businesses. Overall, while Excel or BI tools can be useful for small businesses with simpler spending needs, machine learning provides a more accurate, efficient, and scalable solution for spend classification.



Supplier	Spend (EUR)	Category L1	Category L2	Organization L1	Legal company	Purchasing currency	Invoice count
1 Sky	8,288	Indirect	Logistics	EMEA	Brexit (D78)	EUR	3
1Worldsync	8,250	Indirect	HR & Benefits	Americas	Crunchie Inc (2154)	USD	1
1Worldsync	153,396	Indirect	IT & Telecom	Americas	Crunchie Foods (2132)	USD	4
1Worldsync	6,245	Indirect	IT & Telecom	Americas	Crunchie Inc (2154)	USD	5
1Worldsync	3,997	Indirect	Logistics	Americas	Crunchie Foods (2132)	USD	1
2 AST	3,715	Indirect	Professional Services	Americas	Crunchie Foods (2132)	USD	3
28	660,971	Indirect	Marketing & Communication	Americas	Crunchie Foods (2132)	USD	64
25	91,109	Indirect	Facility Management	Americas	Crunchie Global (2259)	BRL	37
25	38,571	Indirect	IT & Telecom	Americas	Crunchie Global (2259)	BRL	15
11 3C	172,486	Indirect	Professional Services	Americas	Crunchie Foods (2132)	USD	10
12 3C	3,561	Indirect	Professional Services	Americas	Crunchie Inc (2154)	USD	2
13 3D Modelagem e Comercio de Embalagens	1,573	Indirect	Professional Services	Americas	Crunchie Global (2259)	BRL	2
14 3F	1,553	Indirect	Professional Services	Americas	Crunchie Global (2259)	BRL	1
15 3M Company	207,468	Indirect	Facility Management	EMEA	Krinity Crab Sea (420)	NOK	343
16 3M Company	57,451	Indirect	MRO	Americas	Crunchie Canada (2133)	CAD	13
17 3M Company	364,801	Indirect	MRO	Americas	Crunchie Inc (2154)	USD	139
18 3PL Solutions	2,215,373	Indirect	Logistics	Americas	Crunchie Inc (2154)	USD	49
4 Star Hose and Supply	89,498	Indirect	MRO	Americas	Crunchie Inc (2154)	USD	126
4-11v Agencia de Propaganda e Publidade	15,325	Indirect	Professional Services	Americas	Crunchie Global (2259)	BRL	3
4teprint Group	498	Indirect	Professional Services	Americas	Crunchie Inc (2154)	USD	1
45light	9,688	Indirect	Marketing & Communication	Americas	Crunchie Inc (2154)	USD	1
54ra Inovacao Digital	700	Indirect	Professional Services	Americas	Crunchie Global (2259)	BRL	3
07 Oslo Olsen	709,351	Indirect	Marketing & Communication	EMEA	Krinity Crab Sea (420)	NOK	509
07 Tross	650,636	Indirect	Marketing & Communication	EMEA	Krinity Crab Sea (420)	NOK	3,184
7-Eleven Canada	1,540	Indirect	Marketing & Communication	Americas	Crunchie Canada (2133)	CAD	1
27 Tabeliao de Notas Capital	18,076	Indirect	Marketing & Communication	Americas	Crunchie Global (2259)	BRL	10
48Porty Solutions	4,537	Indirect	Logistics	Americas	Crunchie Foods (2132)	USD	6
48Porty Solutions	4,614,416	Indirect	Logistics	Americas	Crunchie Inc (2154)	USD	342

Image: Analyzing the spend data in Excel. Source:sievo.com

In the context of Procurement, machine learning techniques can be utilized to classify spend more accurately or efficiently than data classified by human practitioners alone.



Example of unsupervised learning – vendor matching. Source: sievo.com

3.0 Target Specifications and characterization

The model should be able to analyze financial data from different sources such as credit card statements, bank statements, and receipts, and categorize the expenditures into relevant categories such as office supplies, rent, utilities, inventory, and employee expenses. The goal is to provide small businesses with a better understanding of their spending habits, identify potential areas for cost savings, and optimize their budget allocation to improve their financial performance. Small businesses often have limited resources and need to be more efficient in managing their expenses, and spend classification can provide valuable insights to achieve this goal.

4.0 External Search (online references/links)

- [Machine Learning \(ML\) and Spend Classification \(strategypeeps.com\)](https://www.strategypeeps.com/machine-learning/ml-and-spend-classification)
- <https://www.mckinsey.com/capabilities/operations/how-we-help-clients/product-development-procurement/actionable-spend-insights-orpheus>
- <https://www.sdcexec.com/sourcing-procurement/news/21710545/fairmarkit-tail-spend-becomes-increasing-focus-for-companies>
- <https://www.pwc.in/pwcs-technology-solutions/finance-transformation/pwc-business-analytics-and-insights-platform.html>
- <https://sievo.com/resources/spend-analysis-101>
- <https://www.cl.uni-heidelberg.de/~mroth/pub/spend-class-08.pdf>
- <https://sievo.com/resources/webinars/best-practices-of-implementing-spend-analysis>
- https://get.coupa.com/rs/950-OLU-185/images/Coupa_DataMap_Procurement_Transformation_eBook.pdf
- https://www.researchgate.net/publication/355807922_AgileML_A_Machine_Learning_Project_Development_Pipeline_Incorporating_Active_Consumer_Engagement

5.0 Benchmarking alternate products

- AppZen
- Sievo
- Datamolino
- Coupa
- Procurify
- SpendHQ

These services use machine learning and AI algorithms to automatically classify spending data, provide real-time insights into spending patterns and supplier performance, and detect anomalies or potential fraud. Each solution may have different features, pricing, and integration options, so the pricing isn't publicly available. We need to make a new service from scratch catering to the needs of a specific company.

6.0 Applicable Patents

- Patent US8239335B2: Data classification using machine learning techniques
Current Assignee: Kofax Inc
<https://patents.google.com/patent/US8239335B2/en>
Anticipated expiration: 23.05.2027

7.0 Applicable Regulations

- The use of spend classification software or services may be subject to the Information Technology (IT) Act, 2000, which governs electronic transactions and data privacy.
- Personal Data Protection Bill, 2022. The bill seeks to regulate the collection, storage, and processing of personal data and imposes penalties for non-compliance.
- Labour laws.
- Laws pertaining to tax regulations.

8.0 Applicable Constraints

- **Limited resources:** Small businesses may have limited financial and human resources to invest in implementing spend classification software or services. This can make it difficult to justify the cost of implementing such a system. Use of cloud platform and converting the physical invoice/bills to digital data is a constraint
- **Data quality:** Spend classification relies on accurate and complete data to be effective. Small businesses may not have the resources to maintain high-quality data, which can lead to inaccurate classifications and limited insights. Also, this service assumes all the transactions data are digitally available.
For example, a new computer may be labeled as IT equipment in the general ledger, while the invoice line description provides additional detail distinguishing it as a desktop computer. The purchase order for this item may even have a different description, referring to a vendor or maker-specific data points. While all these data sources refer to the same item, it requires intelligence to make a correct classification.
- **Integration with existing systems:** Spend classification software or services may need to be integrated with existing accounting or procurement systems. This can be challenging for small businesses that have limited IT resources or outdated systems.
- **Resistance to change:** Small businesses may be resistant to change or hesitant to adopt new technologies, especially if they have been using manual processes for a long time.

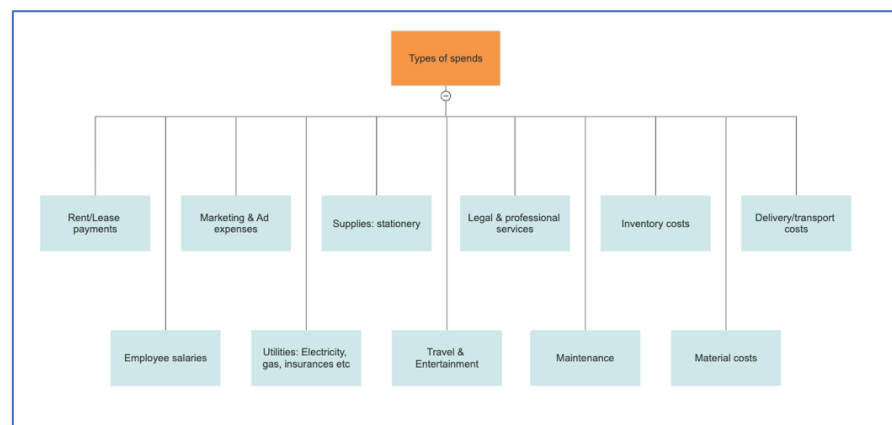
9.0 Business Model

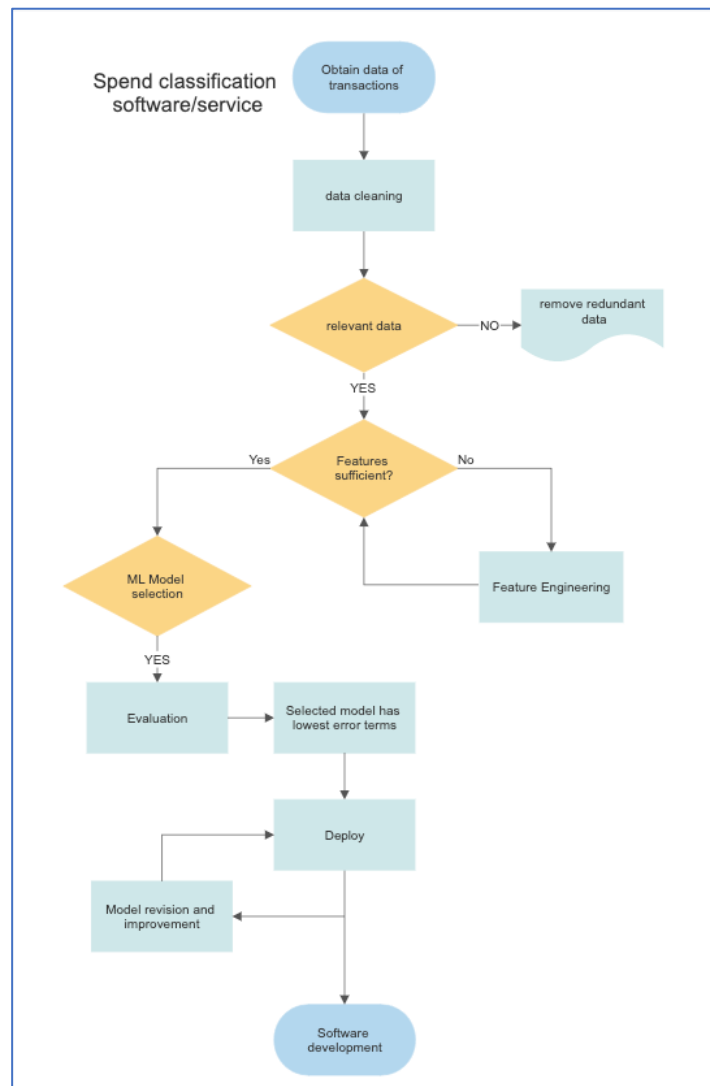
here is a possible business model for a spend classification software or service provider targeting small businesses:

- **Revenue Streams:** Subscription fees for the use of the spend classification software or service.
- **Customer Segments:** Small businesses in various industries that need a cost-effective solution for automating spend analysis.
- **Value Proposition:** Save time and money by automating the classification of expenses, providing insights into spending patterns, and identifying potential cost savings.
- **Channels:** Social media, targeted advertising, email marketing, and partnerships with accounting or procurement firms.
- **Key Activities:** Software development and maintenance, customer support, and marketing and sales.
- **Key Resources:** Skilled software developers, customer support staff, sales and marketing team, infrastructure and technology.
- **Cost Structure:** Balance costs with revenue generated from subscriptions to ensure profitability.

Overall, this business model would focus on providing a cost-effective and easy-to-use solution for small businesses to automate their spend analysis process by doing SWOT analyses and competitive strategy, while also ensuring that the software is continuously updated and improved to meet customer needs.

10.0 Concept Generation





11.0 Concept Development

Machine learning (ML) algorithms can classify spend by learning patterns and relationships in data. These algorithms are trained on historical data that is labeled with expense categories, such as office supplies, travel expenses, or marketing costs. The ML model then uses these labeled examples to learn how to classify new, unlabeled data accurately.

There are various ML algorithms that can be used for spend classification, such as decision trees, random forests, and neural networks. These algorithms work by identifying features or attributes in the expense data that are most relevant to predicting the expense category. For example, the algorithm may look at the expense amount, vendor name, or payment method to determine the category.

Once the ML model is trained, it can be used to classify new, unlabeled expenses automatically. The model takes in the new expense data and outputs the predicted expense category. The accuracy of the model depends on the quality and quantity of the labeled training data and the complexity of the algorithm used. With accurate spend classification, businesses can gain insights into their spending patterns and identify areas for cost savings or optimization.

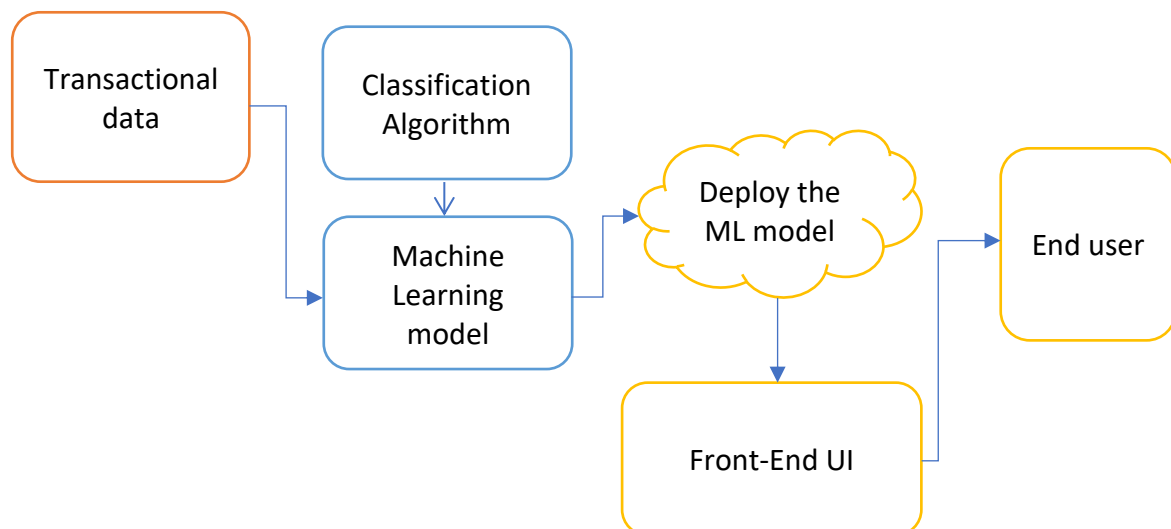
There are few ML/AI algorithms used for spend classification:

- **Decision Trees:** Decision trees are a popular algorithm for spend classification because they are easy to interpret and visualize. They work by splitting the data into smaller and smaller subsets based on certain criteria, eventually leading to a final prediction or classification.
- **Random Forests:** Random forests are an ensemble learning method that combine multiple decision trees to improve accuracy and reduce overfitting. They are a popular choice for spend classification because they can handle large amounts of data with high dimensionality.
- **Naive Bayes:** Naive Bayes is a probabilistic algorithm that is often used for text classification, but can also be used for spend classification. It works by calculating the probability of each spending category based on the features of the data.
- **Support Vector Machines (SVMs):** SVMs are a popular algorithm for classification problems, including spend classification. They work by finding the optimal hyperplane that separates the data into different classes.
- **Neural Networks:** Neural networks are a powerful algorithm that can be used for spend classification, especially when dealing with large amounts of data with high dimensionality. They are a popular choice for deep learning and can handle complex relationships within the data.

It's always recommended to test and compare different algorithms to find the best fit for the problem at hand.

Using the cleaned spend data and classifying it using above mentioned ML algorithms. And Utilizing one algorithm with least errors. This concept can be developed using the API or integration with BI tool where our ML algorithm can classify the data with higher accuracy.

12.0 Final Product Prototype



13.0 Product Details

It is a service/software which takes transactional data, finds patterns, trains itself using that data and gives an outcome. ML algorithms has many advantages over conventional analysis softwares.

- **Scalability:** As the volume of data grows, it can become challenging to manage and analyze with traditional BI tools. Machine learning can scale more effectively, handling large volumes of data without impacting performance.
- **Complexity:** Financial data can be complex, with multiple variables and factors that can impact spending patterns. Machine learning can identify these complex relationships and provide more accurate insights into spending behavior.
- **Accuracy:** Machine learning models can learn from data and improve over time, increasing accuracy and reducing errors. BI tools can provide static reports, but machine learning can continually analyze new data and update spending patterns, providing more accurate insights.
- **Flexibility:** Machine learning models can be trained on different types of data, making them more flexible than BI tools. They can adapt to changes in spending patterns and adjust spending categorization accordingly.

13.1 Algorithms used

As mentioned above, ML models are used and based on the evaluation metrics, the best performing model is used for spend classification.

13.2 Team required to develop

- ML Engineer
- Financial Analyst
- Software developer
- Cloud Engineer

14.0 Conclusion

Spend classification ML models are primarily used to classify expenses into different categories based on their attributes such as description, amount, date, and other factors. These models do not directly help with prediction as they do not provide insights into future trends or spending patterns.

However, spend classification can be a critical first step in enabling predictive analytics models for small businesses. By categorizing historical expenses accurately, businesses can use this data to develop predictive models for forecasting future expenses, identifying potential cost savings opportunities, and optimizing their spending decisions.

For example, a business might use historical spend data to develop a predictive model that forecasts future expenses for a particular category. The model could then be used to identify areas where expenses are likely to increase or decrease in the future, allowing the business to adjust its spending priorities accordingly.

15.0 References

- [1] Zhang, D., Yang, Z., & Liu, Y. (2020). An intelligent spend classification model for small businesses based on machine learning. *Journal of Intelligent & Fuzzy Systems*, 39(1), 1319-1330.

- [2] Siau, K., Wang, W., & Sheng, H. (2021). An intelligent spend classification framework for small and medium-sized enterprises. *Journal of Intelligent & Fuzzy Systems*, 41(2), 2177-2189.
- [3] Lee, J. K., & Song, Y. J. (2019). Development of a spend classification model for small and medium-sized enterprises using machine learning techniques. *Journal of Industrial Distribution & Business*, 10(1), 5-11.
- [4] Wang, X., Zhang, Q., & Yao, X. (2021). A classification model for small business spend based on machine learning. *IEEE Access*, 9, 139338-139349.
- [5] Cao, J., Liu, M., & Li, X. (2020). An intelligent spend classification system for small and medium-sized enterprises. In *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 976-980). IEEE.