

Vitor Augusto Menten de Barros

Vision Palette: Use of Artificial Intelligence for Personal Color Classification

SÃO PAULO
2025

Vitor Augusto Menten de Barros

Vision Palette: Use of Artificial Intelligence for Personal Color Classification

Final Course Project submitted to the
Institute of Technology and Leadership
(INTELI), to obtain a bachelor's degree in
Computer Science

Advisor: Prof. Tomaz Mikio Sasaki

SÃO PAULO
2025

Cataloging in Publication
Library and Documentation Service
Institute of Technology and Leadership (INTELLI)
Data entered by the author.

(Cataloging record with international cataloging data, according to NBR 14724. The record will be completed later, after approval and before the final version is deposited. The completion of the cataloging record is the responsibility of the institution's library.)

Sobrenome, Nome

Título do trabalho: subtítulo / Nome Sobrenome do autor; Nome e Sobrenome do orientador. – São Paulo, 2025.
nº de páginas : il.

Trabalho de Conclusão de Curso (Graduação) – Curso de [Ciência da Computação] [Engenharia de Software] [Engenharia de Hardware] [Sistema de Informação] / Instituto de Tecnologia e Liderança.

Bibliografia

1. [Assunto A]. 2. [Assunto B]. 3. [Assunto C].

CDD. 23. ed.

Acknowledgments

I would like to express my sincere gratitude to Andreia Menten da Silva Barros, partner at Experimenten, for her valuable contributions throughout the development of this project. Her continuous feedback, detailed explanations of the personal color analysis process, and insights into the practical pain points of the consultancy were fundamental to understanding the business context and guiding the development of the proposed solution. Her support and domain expertise greatly enriched both the technical and applied aspects of this work.

Resumo

Barros, Vitor. **Vision Palette: Uso de Inteligência Artificial para a Classificação de Coloração Pessoal.** 2025. 48. TCC (Graduação) – Curso Ciência da Computação, Instituto de Tecnologia e Liderança, São Paulo, 2025.

Este trabalho apresenta o desenvolvimento e a avaliação de uma solução baseada em inteligência artificial para a classificação de coloração pessoal, realizada em parceria com uma empresa do setor de moda e acessórios que integra consultoria de imagem aos seus serviços. A análise de coloração pessoal é tradicionalmente conduzida por meio de um processo manual, demorado e fortemente dependente da experiência profissional, o que limita sua escalabilidade e eleva os custos operacionais. O objetivo inicial do projeto foi desenvolver uma aplicação totalmente automatizada, pronta para uso pelo usuário final, capaz de determinar a paleta de cores pessoais de um indivíduo sem a necessidade de acompanhamento profissional. Para isso, foi projetado um sistema web utilizando uma arquitetura de microserviços em nuvem, combinando um frontend em Next.js, um backend em NestJS e um serviço de inferência em visão computacional implementado com FastAPI e TensorFlow. Foram exploradas diversas arquiteturas de Redes Neurais Convolucionais, incluindo modelos rasos treinados do zero, técnicas de transferência de aprendizado com redes pré-treinadas e estratégias alternativas de modelagem, como classificação hierárquica, classificação em menor granularidade e uma abordagem multi-modelo baseada em atributos perceptuais. O desempenho dos modelos foi avaliado por meio de conjuntos de dados sintéticos controlados e conjuntos de validação com imagens do mundo real, apresentando diferentes condições de qualidade e ambiente. Os resultados demonstram que a classificação de alta granularidade em doze paletas sazonais é altamente sensível às limitações do conjunto de dados e à variabilidade do mundo real, impedindo o alcance completo do objetivo inicial de automação total. Ainda assim, a solução proposta mostrou-se eficaz como uma ferramenta auxiliar de apoio à decisão, oferecendo inferência em tempo quase real, com tempos de resposta inferiores a um segundo, e possibilitando reduções na etapa inicial de análise superiores à meta de 75%. Do ponto de vista do negócio, a solução apresenta-se como um serviço digital escalável e de baixo custo, capaz de viabilizar fluxos híbridos que combinam pré-classificação automatizada com interpretação profissional. O trabalho evidencia a importância crítica da representatividade dos dados em sistemas de visão computacional aplicados e estabelece uma base sólida para evoluções futuras voltadas ao aumento da robustez, da precisão e da viabilidade prática da solução.

Palavras-Chave: coloração pessoal; visão computacional; redes neurais convolucionais; aprendizado de máquina; tecnologia da moda.

ABSTRACT

Vitor, Barros. **Vision Palette: Use of Artificial Intelligence for Personal Color Classification.** 2025. 48. Final course project (Bachelor) – Course Computer Science, Institute of Technology and Leadership, São Paulo, 2025.

This work presents the development and evaluation of an artificial intelligence-based solution for personal color classification, conducted in partnership with a fashion and accessories company that integrates image consultancy into its services. Personal color analysis is traditionally performed through a manual, time-intensive, and expertise-dependent process, which limits scalability and increases operational cost. The initial objective of the project was to develop a fully automated, user-ready application capable of determining an individual's personal color palette without professional consultancy. To achieve this, a web-based system was designed using a cloud-based microservices architecture, combining a Next.js frontend, a NestJS backend, and a computer vision inference service implemented with FastAPI and TensorFlow. Several Convolutional Neural Network (CNN) architectures were explored, including shallow models trained from scratch, transfer learning with pre-trained networks, and alternative modeling strategies such as hierarchical gating, coarse-grained classification, and a multi-model decomposition based on perceptual attributes. Model performance was evaluated using controlled synthetic datasets and real-world validation sets with varying image quality and environmental conditions. The results demonstrate that high-granularity classification into twelve seasonal palettes is highly sensitive to dataset limitations and real-world variability, preventing the full achievement of the initial automation objective. However, the proposed solution proved effective as an auxiliary decision-support tool, delivering near real-time inference with sub-second response times and enabling reductions in the initial analysis stage well beyond the targeted 75%. From a business perspective, the solution offers a scalable and low-cost digital service capable of supporting hybrid workflows that combine automated pre-classification with professional interpretation. The project highlights the critical importance of dataset representativeness for applied computer vision systems and establishes a clear foundation for future improvements aimed at increasing robustness, accuracy, and practical viability.

Key words: personal color analysis; computer vision; convolutional neural networks; machine learning; fashion technology.

List of Illustrations

Figure 1 - User image and its corresponding palette..	pág. 13
Figure 2 – Technology Stack.....	pág. 19
Figure 3 – Deployment Diagram.....	pág. 28
Figure 4 – Backend Unit Tests Results.....	pág. 33
Figure 5 – Frontend Unit Tests Results.....	pág. 34
Figure 6 – E2E Tests Results.....	pág. 35

List of Tables

Table 1 – Accuracy Results for Each Approach.....	pág. 39
Table 2 – Inference and Response Time for Each Model.....	pág. 40

List of Abbreviations and Acronyms

CNN – Convolutional Neural Network
ORM – Object-Relational Mapping
E2E – End-to-End
MVP – Minimum Viable Product
AWS – Amazon Web Services
UAT – User Acceptance Testing
KPIs – Key Performance Indicators

Summary

1 Introduction	9
2 Solution Development	15
2.1 Applied Rationale	15
2.2 Specification and Development:	21
2.3 Assessment of Impact and Contribution to the Business	34
3 Conclusion	43
References	45

1 Introduction

1.1. Partner Company Context:

Experimenten is a small-to-medium-sized enterprise operating in the fashion and accessories sector, with a specific focus on semi-jewelry. The company combines aesthetic curation with principles of personal coloration and image consulting, offering both a physical retail space in São Paulo and a structured e-commerce platform. Its target audience consists primarily of women interested in accessories that not only complement their style but also reinforce self-perception, personal identity, and visual coherence.

Although operating at a boutique scale, Experimenten maintains a diversified internal structure that includes product curation, digital marketing, customer experience, and a dedicated Personal Coloration and Styling division. This division has strategic relevance within the organization, as it is responsible for guiding customers toward accessories aligned with their natural chromatic characteristics, producing educational content related to image consulting, and supporting the brand's positioning as a reference in style consultancy integrated with fashion. Consequently, this division represents the primary area affected by the present project, directly influencing aesthetic direction, customer experience standards, and communication strategies across both physical and digital channels.

Technni, the technological partner in this project, operates in the software development and cloud infrastructure sector. The company specializes in the design and implementation of digital products, backend and frontend systems, database management, and the integration of artificial intelligence into production environments. Within the scope of this partnership, Technni is responsible for translating Experimenten's business requirements into scalable technological solutions, including the development of machine learning pipelines, computer vision services, and cloud-based deployment architectures. These activities align with Technni's broader role in supporting Experimenten's technological roadmap and enabling the operationalization of data-driven tools.

The partnership between Experimenten and Technni is strategically relevant for both organizations. For Experimenten, the integration of structured personal color analysis methodologies with artificial intelligence represents an opportunity to differentiate itself within a highly competitive semi-jewelry market. By systematizing a process that is traditionally manual and experience-dependent, the company can enhance consistency, reduce service time, and expand the reach of its color-analysis services. In addition, the availability of AI-assisted tools supports more informed decisions in product curation, inventory planning, and content creation, reinforcing the brand's positioning and long-term innovation strategy.

From Technni's perspective, the project provides a concrete opportunity to validate and refine its artificial intelligence and cloud infrastructure solutions within a real commercial context. The development of computer vision systems applied to personal coloration allows Technni to expand its portfolio in the fashion-tech domain, demonstrate its capacity to deliver high-complexity digital products, and establish reusable technological frameworks that can be adapted to other sectors. This collaboration also strengthens the strategic relationship between the two companies, fostering knowledge exchange and technological maturation.

Together, Experimenten and Technni form a collaborative ecosystem that integrates domain-specific aesthetic expertise with technological innovation. The project strengthens Experimenten's value proposition, enriches Technni's technological capabilities, and positions both organizations at the forefront of emerging applications of artificial intelligence in personal coloration, styling, and digital fashion experiences. This alignment between business strategy and technological development underscores the relevance of the partnership and supports the academic significance of the present work.

1.2. Problem Definition (Corporate Pain Point):

Personal color analysis is a structured methodology used to evaluate an individual's natural chromatic characteristics—such as skin tone, hair color, and eye color—in order to identify a color palette that harmonizes with and enhances their appearance. Beyond its aesthetic dimension, personal color analysis supports image construction,

facilitates clothing and accessory choices, and influences communication, presence, and self-perception. In this context, the process aims to enhance natural beauty in a balanced and coherent manner, respecting individual characteristics rather than imposing rigid standards (BROWN; ROJAS, 2025).

Within the partner company's current operational model, personal color analysis is conducted entirely through a manual and consultant-dependent process. The analysis involves sequential stages, including contrast evaluation, undertone identification, intensity assessment, and depth analysis, all of which rely on visual inspection, professional judgment, and comparison with physical reference palettes. This process is time-intensive and requires sustained attention and expertise from the consultant throughout the session.

As a result, the average duration of a personal color analysis session is approximately 120 minutes, including preparation, analysis, explanation of results, and guidance on practical application. This duration directly limits service throughput, typically restricting consultants to a small number of sessions per day. Consequently, the service presents high operational costs, both in terms of labor and opportunity cost, as consultant time could otherwise be allocated to more scalable or higher-margin services and products offered by the company.

In addition to time constraints, the manual nature of the process introduces variability and susceptibility to error. The accuracy of the analysis is highly dependent on the consultant's level of experience, visual perception, and environmental conditions such as lighting. Less experienced professionals may reach inconsistent conclusions even when following the same methodological framework, increasing the risk of classification errors and reducing standardization across sessions.

From the user's perspective, this operational model also presents significant limitations. Clients are required to invest a considerable amount of time in a single session, which may be perceived as tiring or inconvenient. The service is typically associated with a high monetary cost, reflecting its duration and reliance on specialized expertise. In many cases, users must also travel to the consultant's physical location, further increasing time and financial expenditure and reducing accessibility.

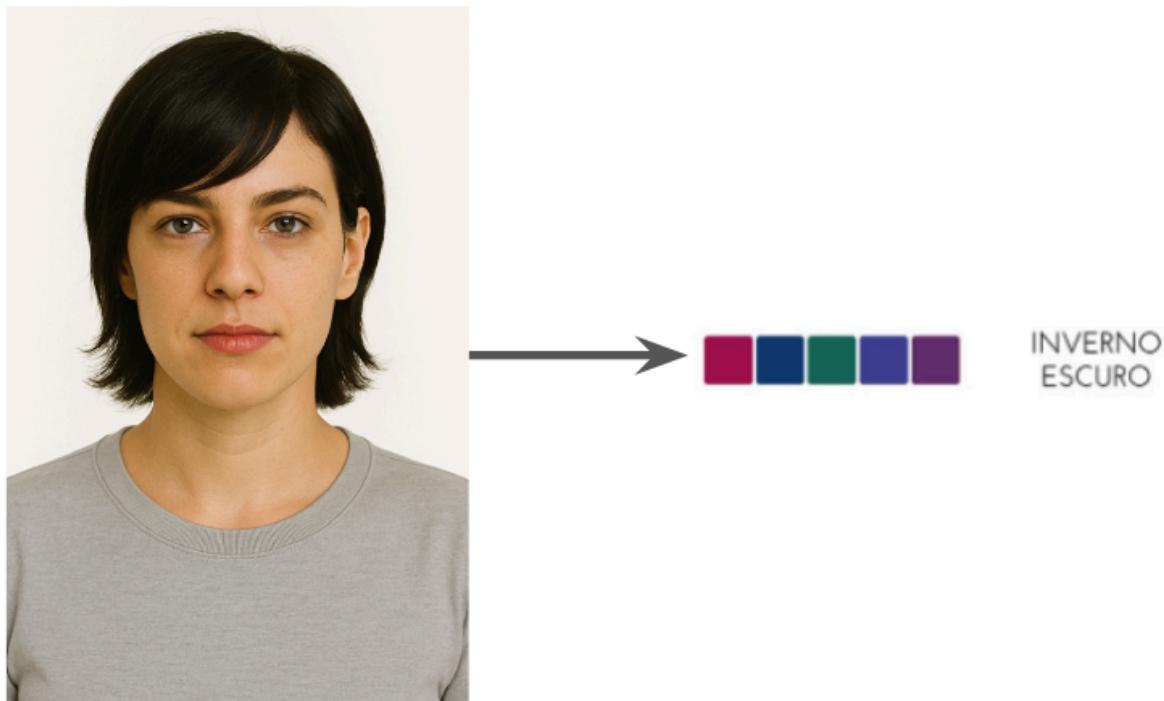
Taken together, the current baseline metrics indicate a process characterized by long session duration (\approx 120 minutes per analysis), limited scalability, high dependency on human expertise, and elevated operational and user costs. These factors constrain the company's ability to expand its personal color services and limit their integration into broader digital offerings.

In response to these challenges, the present project seeks to address the identified pain points by reducing average session time, minimizing subjectivity and error rates, and lowering operational costs for both consultants and users through the application of artificial intelligence and computer vision techniques.

1.3. Proposed Solution and Expected Contribution:

The proposed solution is a web-based software tool designed to support the personal color analysis process through the application of computer vision and artificial intelligence techniques. The system enables users to upload a facial photograph via a web interface, after which the image is processed by a machine learning model trained to analyze visual features related to skin tone, hair color, and eye color. Based on this analysis, the tool returns a suggested personal color palette aligned with the user's natural chromatic characteristics. Figure 1 illustrates an example of a user-provided input image together with the corresponding color palette returned by the system after prediction.

Figure 1 - User image and its corresponding palette



Source: Prepared by the author.

The system operates on a fixed payment model: one image, one payment, making the service more accessible and straightforward. Users will pay for each analysis, enabling a consistent, scalable revenue stream for the business and simplifying the user's financial commitment.

In addition to automated classification, the platform provides contextual and educational content explaining the principles of personal color analysis and the meaning of the suggested palette. This component is intended to support user understanding, promote transparency in the decision-making process, and position the tool as an assistive resource rather than a purely opaque automated system.

From a computational perspective, the solution is built around a Convolutional Neural Network (CNN) architecture, selected due to its established effectiveness in image-based pattern recognition tasks (XU et al., 2024). Model training was conducted using a dataset composed of both artificially generated images—created to follow standardized visual constraints—and real-world images of individuals whose personal color palettes had been previously determined by professional consultants. This hybrid dataset strategy was adopted to balance controlled learning conditions with exposure to realistic visual variability.

The primary expected contribution of the project is the significant reduction of personal color analysis session time. Based on the current baseline of approximately 120 minutes per session, the project establishes a quantitative objective of achieving a time reduction of at least 75%, enabling the delivery of an initial analysis result in a matter of minutes. This reduction directly addresses scalability, operational cost, and accessibility constraints identified in the problem definition.

While the evaluation of accuracy was framed as a key project metric, the main goal was to develop a tool capable of serving as a practical replacement for the manual, consultant-dependent process. The system was designed to provide users with the ability to generate their own color palette autonomously. Even though the goal was to fully replace the consultant in the analysis process, the accuracy benchmark was set at approximately 85% to evaluate performance rather than being an initial business expectation. The aim was to empirically assess the feasibility of the system, with the understanding that the tool would be an assistive decision-support system capable of guiding users in color palette selection without requiring professional intervention.

Overall, the proposed solution aims to function as a replacement for the traditional process, allowing users to independently obtain their personalized color palette quickly and cost-effectively, thereby reducing reliance on human consultants and increasing the scalability of the company's color analysis services.

1.4. Business Objectives:

The expected results of this project for the partner company are primarily related to scalability, operational efficiency, revenue potential, and strategic positioning. By introducing an AI-based personal color analysis tool, Experimenten is expected to expand its service capacity without a proportional increase in operational costs or dependence on specialized human resources.

From an operational perspective, the reduction in analysis time enables a higher volume of analyses to be delivered within the same time frame, allowing the company to serve a larger number of users. This increase in throughput directly supports revenue growth, particularly under the proposed fixed payment model (one image per analysis), which enables a standardized and scalable monetization

strategy. In addition, by automating the initial stages of personal color analysis, consultants can reallocate their time to higher-value activities, such as personalized styling services, premium consultations, content creation, and product curation.

The project is also expected to positively impact the company's profitability. Automation reduces the marginal cost per analysis, enabling Experimenten to offer personal color services at a lower price point while maintaining or improving profit margins. This cost structure supports the expansion of the service to a broader audience, increasing market reach and customer acquisition.

From a strategic and marketing standpoint, the implementation of an AI-assisted color analysis tool positions Experimenten as an innovative brand within the semi-jewelry and image consulting market. The availability of a digital, technology-driven solution differentiates the company from competitors that rely exclusively on manual processes, strengthening brand perception and reinforcing its positioning at the intersection of fashion, consulting, and technology. This differentiation is expected to increase engagement with the company's digital channels and to act as a gateway for cross-selling products and complementary services.

User-level benefits play an important indirect role in achieving these results. Increased confidence in color choices, improved self-knowledge, time and cost savings, and a more optimized wardrobe contribute to higher user satisfaction and perceived value. These factors are expected to improve conversion rates, customer retention, and long-term relationship building, thereby reinforcing the company-level benefits outlined above.

Overall, the project aims to deliver measurable business value by enabling Experimenten to scale its personal color services, improve operational efficiency, strengthen its market positioning, and support sustainable growth through the integration of artificial intelligence into its service portfolio.

1.5. Structure of the thesis/dissertation:

This thesis is structured into three main chapters. Chapter 1 introduces the project by presenting the partner company context, the problem definition, the proposed solution, and the expected business contributions, establishing the motivation and objectives of the work. Chapter 2 details the development of the proposed solution, including the applied business and technological rationale, system specifications, architectural design, implementation process, and technical evaluation of the system. Finally, Chapter 3 presents the conclusions of the project, discussing the achieved results, limitations, lessons learned, and recommendations for future work and system evolution.

2 Solution Development

2.1 Applied Rationale

2.1.1 Business Area Rationale:

Personal color analysis is a structured technique used to identify color palettes that harmonize with an individual's natural chromatic characteristics, such as skin tone, hair color, and eye color. Rather than imposing prescriptive rules, its purpose is to guide individuals toward color choices that enhance appearance, promote visual balance, and strengthen personal image. Within the partner company's sector, personal color analysis is regarded not only as an aesthetic tool but also as a strategic resource that supports image construction, communication, and consumer confidence.

Colors that align with an individual's natural features tend to illuminate the face, reduce the appearance of visual fatigue or imperfections, and emphasize facial expression. Conversely, poorly aligned colors may create shadows, diminish skin brightness, and accentuate features that would otherwise be less noticeable. As a result, color selection plays a significant role in how personal image is perceived, both by the individual and by others. From a practical standpoint, personal color analysis simplifies decision-making related to clothing, accessories, and makeup, reducing uncertainty and minimizing impulsive purchasing behavior.

Within the partner company's operational context, personal color analysis is conducted through a sequence of well-defined analytical stages. These stages provide a structured framework for evaluating chromatic characteristics and are summarized as follows.

The contrast analysis stage evaluates the level of contrast between natural features such as skin, eyes, and hair, typically classified as low, medium, or high. This dimension determines whether an individual is better complemented by soft, balanced, or bold color combinations. Contrast is considered a foundational element, as visual harmony is achieved by reflecting an individual's natural contrast level in clothing, accessories, makeup, and hair color.

The temperature (undertone) analysis stage identifies whether an individual's undertone is warm or cool. Warm undertones are commonly associated with hues such as yellow, beige, coral, warm brown, or peach, while cool undertones may present pinkish, bluish, porcelain, olive, or neutral characteristics. This assessment guides the selection of color temperatures that maintain visual balance and avoid chromatic disharmony.

The intensity analysis stage evaluates whether an individual's features are better enhanced by bright, vivid colors with high chromatic purity or by softer, muted tones influenced by gray or brown pigments. This dimension is essential for defining whether a visual identity should emphasize vibrancy or subtlety.

Finally, the depth analysis stage examines whether lighter or darker colors create greater harmony when combined with an individual's skin tone, hair color, and eye color. This assessment defines the most suitable range of lightness or darkness for achieving a balanced appearance.

Together, these four dimensions—contrast, temperature, intensity, and depth—form the conceptual foundation of personal color analysis within the partner company's sector. Building upon this foundation, the Expanded Seasonal Method provides a more granular classification framework. This method extends the traditional four-season model (Spring, Summer, Autumn, and Winter) into twelve distinct categories, enabling more precise and personalized color recommendations. Each classification results from the combined interpretation of the four analytical

dimensions as they manifest in an individual's physical characteristics, making the method particularly suitable for structured labeling and classification tasks.

From a market perspective, best practices in personal color analysis remain predominantly manual and expertise-driven. Professional consultants rely on visual assessment, structured analytical methodologies, and physical reference tools—such as color drapes and standardized palettes—typically applied in controlled environments with neutral lighting. Beyond the analytical session itself, the value of personal color consultancy is not limited to the one or two hours dedicated to testing, but rather encompasses a broader set of deliverables designed for long-term use.

A complete personal color consultancy package commonly includes an individualized diagnostic session, a physical color palette card intended to guide purchasing decisions, a personalized digital dossier explaining the identified palette and its applications to clothing, makeup, accessories, and hair. These complementary materials reinforce the perceived value of the service and position personal color analysis as a long-term investment rather than a single diagnostic interaction.

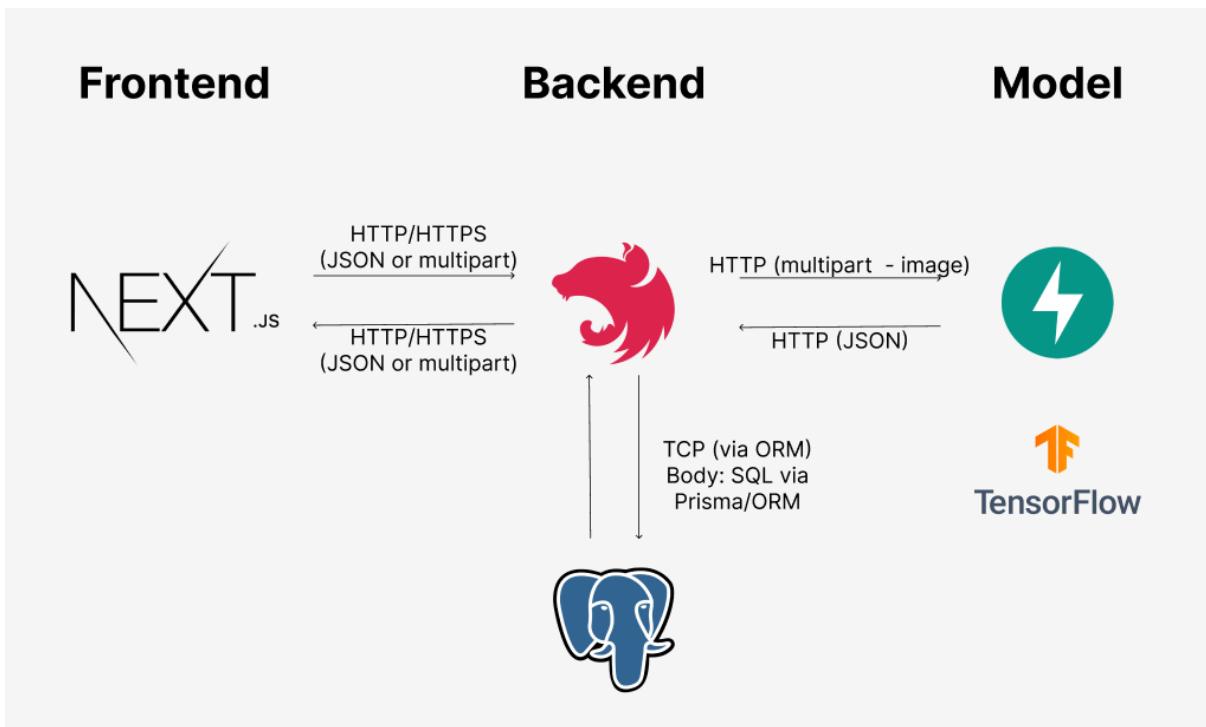
Benchmarking within the sector indicates that the quality and consistency of outcomes are more strongly correlated with professional expertise than with the analytical tools employed. Experienced consultants tend to deliver more reliable and coherent classifications, while less experienced practitioners may reach divergent conclusions even when applying the same methodological framework. Consequently, accumulated experience, perceptual training, and learning curves remain key differentiators in service quality.

Within this context, the process addressed in this project aligns with established market practices while revealing a relevant structural limitation: the heavy reliance on human perception and individual expertise as the primary determinants of accuracy, consistency, and scalability. This limitation highlights an opportunity for technological support solutions capable of standardizing and accelerating the analytical stage of personal color assessment, while preserving the conceptual foundations and complementary value elements already validated within the professional consultancy market.

2.1.2 Technological rationale for the solution:

2.1.2.1 Services Development

Figure 2 – Technology Stack



Source: Prepared by the author.

Figure 1 presents the overall system architecture and technology stack adopted in this project. The technological choices were guided by the partner company's operational context, which includes a cloud-based infrastructure, limited computational resources typical of small-to-medium-sized enterprises, and the need for scalability, maintainability, and efficient deployment of machine learning services.

The frontend layer is responsible for the graphical user interface through which users interact with the system. It was implemented using Next.js, a modern framework built on React that supports Server-Side Rendering and Static Site Generation. These features were considered particularly suitable for a user-facing product, as they improve initial load performance, accessibility, and search engine optimization, while maintaining flexibility for dynamic content and future expansion (NEXT.JS, 2025). From a business perspective, Next.js supports the development of performant and responsive interfaces without introducing excessive architectural complexity.

The backend layer is responsible for handling client requests, enforcing business rules, managing user data, and orchestrating communication between system components. This layer was implemented using NestJS, a Node.js framework designed around a modular architecture with native support for dependency injection and decorators. These characteristics facilitate code organization, scalability, and long-term maintainability, which are essential for systems expected to evolve beyond a minimum viable product. Additionally, NestJS offers strong integration with relational databases through Object-Relational Mapping (ORM) tools such as Prisma or TypeORM, aligning with the partner company's data persistence requirements (BEIGHTON, 2025).

The computer vision component, referred to as the VisionPalette model service, was implemented as an independent API using FastAPI. This service receives images forwarded by the backend, performs model inference using TensorFlow, and returns classification results in JSON format. FastAPI was selected due to its high performance, native support for asynchronous request handling, and suitability for machine learning inference workloads. Compared to older frameworks such as Flask, FastAPI provides lower latency and improved concurrency, which is particularly relevant for real-time or near real-time prediction services deployed in cloud environments (FASTAPI, 2025). This separation between application logic and model inference also supports modularity and independent scaling of services.

2.1.2.2 CNN Development

From a machine learning perspective, CNNs were selected as the core modeling approach due to their established effectiveness in image-based pattern recognition tasks. CNNs are particularly well suited for learning spatial hierarchies of visual features, such as edges, textures, and color distributions, which are central to the problem of personal color classification.

The development process began with the implementation of a shallow CNN trained from scratch, composed of two convolution–pooling blocks followed by fully connected layers. This baseline architecture was intentionally simple and served a didactic and experimental purpose. By avoiding transfer learning in the initial stage, it allowed for a controlled evaluation of the fundamental convolution–pooling pipeline

and provided a reference point against which more complex models could be compared.

Subsequently, transfer learning was applied using pre-trained architectures such as VGG16 (SIMONYAN; ZISSERMAN, 2014), ResNet50 (WU; ZHONG; LIU, 2017), and EfficientNetB0 (TAN; LE, 2019). Transfer learning leverages models trained on large-scale datasets, such as ImageNet, and adapts their learned representations to a domain-specific task. This approach is particularly advantageous when the target dataset is limited, as it enables faster convergence, reduced training cost, and improved generalization performance (WEISS; KHOSHGOFTAAR; WANG, 2016). Given the practical constraints of dataset size and variability in this project, transfer learning represented a technically and economically viable strategy.

In parallel, multiple modeling strategies were explored to address the inherent complexity and ambiguity of personal color classification. These strategies differed in structure and granularity:

- 12-Class Prediction Model: A single model trained to directly classify inputs into one of the twelve expanded seasonal palettes. This approach represents the most granular formulation of the problem but also the most challenging, due to visual similarity between classes.
- 12-Class Prediction with Gating: A hierarchical architecture in which an initial gating model predicts one of the four main seasons, followed by a secondary classifier responsible for identifying the corresponding subpalette. This strategy aims to reduce classification ambiguity by narrowing the decision space at each stage.
- 4-Class Prediction Model: A simplified approach that predicts only the four main seasons (Spring, Summer, Autumn, Winter), reducing class granularity in favor of improved robustness and generalization.
- Multi-Model Approach: A modular strategy that decomposes the problem into two binary classification tasks—Temperature (Warm vs. Cool) and Intensity (Light vs. Deep). The outputs of these models are subsequently combined to infer the final seasonal classification. This decomposition prioritizes interpretability and aligns more closely with the conceptual foundations of personal color analysis.

Together, these approaches enabled a systematic exploration of trade-offs between model complexity, interpretability, robustness, and practical applicability. This experimentation strategy was essential for identifying architectures that balance technical performance with the operational and business constraints of the partner company.

2.1.3 Fundamentals of Management and Development Methods:

At the time of this project, the partner company did not adopt a formally defined software development or project management methodology, such as Scrum, Kanban, or PMBOK-based frameworks. Development activities were typically conducted in an ad hoc manner, driven by immediate business needs and technical feasibility rather than by standardized process models. This context is common in small-to-medium-sized enterprises, where flexibility and rapid experimentation often take precedence over formal process documentation.

Given this context, the project followed an academic, iterative, and adaptive development approach, aligned with core principles of Agile development, DevOps practices, and fundamental project management concepts. Rather than enforcing a formal methodology, these principles were applied in a pragmatic way to support incremental delivery, continuous validation, and alignment with business objectives.

At a high level, the project was organized into four trimesters, each with a clearly defined primary objective. Each trimester was structured into five sprints of two weeks, enabling regular inspection and adaptation cycles. The end of each trimester, and in particular the final sprint, included a formal presentation and review with the partner company to validate intermediate results, gather feedback, and adjust priorities for subsequent phases.

The first trimester was dedicated to business and user understanding, focusing on problem framing, definition of the value proposition, identification of stakeholders, user analysis, and the establishment of high-level requirements and system architecture.

The second trimester concentrated on application development, encompassing the design and implementation of the web application, both frontend and backend, as

well as initial integration efforts and early deployment to support testing and validation.

The third trimester focused on the development of the convolutional neural network model, including exploratory data analysis, preprocessing, iterative training, performance evaluation through appropriate metrics, and validation.

Finally, the fourth trimester addressed final refinements to both the application and the model, incorporating feedback obtained throughout the project, exploring improvements to the model and the system architecture, and culminating in the final deployment of the complete solution.

The concrete application of these practices throughout the development and deployment of the solution is described in detail in Section 2.2.3.

2.2 Specification and Development:

2.2.1 Requirements and Specifications:

This section presents the functional and non-functional requirements of the system, as well as the main user interactions represented through use cases. These specifications were defined based on the identified business needs and serve as a foundation for the system's design, implementation, and evaluation.

2.2.1.1 Use Cases

Use cases were defined to model the interactions between users and the system, describing how specific goals are achieved through system functionality. Each use case represents a distinct interaction scenario and contributes to clarifying expected system behavior under normal conditions as well as in exceptional situations. To improve robustness and reliability, each use case is structured into a normal flow and a set of extensions that capture alternative or error conditions (ROSENBERG; STEPHENS, 2007).

At the initial stage of the project, five primary use cases were identified. The most relevant use cases are described below.

2.2.1.1.1 Obtain Information About Personal Color Analysis

Actor: Experimenten's Customer

Normal Flow:

1. The actor accesses the application.
2. The system displays the home page containing information about personal color analysis and the system's functionality.
3. The actor reads the presented information.

Extensions:

- 2a. The system presents additional explanatory content, including the benefits of personal color analysis, an overview of the analysis process, and examples of how the system supports color selection for clothing, makeup, and accessories.

2.2.1.1.2 Create an Account

Actor: Experimenten's Customer

Normal Flow:

1. The actor accesses the application.
2. The actor selects the "Sign Up" option.
3. The actor provides name, phone number, email address, and password.
4. The actor confirms the password.
5. The system validates the provided data.
6. The actor provides payment information, including card number, expiration date, security code, and cardholder name.
7. The system validates and stores the payment information.
8. The system creates the actor account and confirms successful registration.

Extensions:

- 5a. If the email address is already registered, the system requests a different email.

- 5b. If the password and confirmation do not match, the system requests re-entry.
- 5c. If the email format is invalid, the system requests correction.
- 5d. If the phone number format is invalid, the system requests correction.
- 7a. If payment validation fails, the system requests a retry.
- 8a. If account creation fails, the system notifies the actor and requests a retry.
- 8b. If account creation succeeds, the system redirects the actor to the home page.

2.2.1.1.3 Login

Actor: Experimenten's Customer

Normal Flow:

1. The actor accesses the application.
2. The actor selects the “Login” option.
3. The actor provides email address and password.
4. The system validates credentials and authenticates the actor.
5. The system confirms successful login and redirects the actor to the home page.

Extensions:

- 3a. The system allows the actor to show or hide the password input.
- 4a. If credentials are invalid, the system requests corrected input.
- 4b. If the email format is invalid, the system requests correction.
- 5a. If authentication fails, the system notifies the actor and requests a retry.

2.2.1.1.4 Receive Color Palette Analysis

Actor: Experimenten's Customer

Normal Flow:

1. The actor accesses the application.
2. The actor selects the “Analysis” option.

3. The system presents guidelines for capturing a valid image, including lighting conditions, facial positioning, and the removal of accessories or makeup.
4. The actor uploads an image.
5. The system processes the image using a computer vision-based model.
6. The system displays the resulting personal color palette and a brief explanation.

Extensions:

- 4a. If the image does not comply with guidelines, the system requests a new upload.
- 5a. If the analysis fails, the system notifies the actor and requests a retry.

2.2.1.1.5 Edit Profile Information

Actor: Experimenten's Customer

Normal Flow:

1. The actor authenticates and accesses the profile section.
2. The system displays current profile information.
3. The actor modifies selected fields.
4. The system validates and stores the updated data.
5. The system confirms the update and redirects the actor to the home page.

Extensions:

- 4a. If the email address is already registered, the system requests a different email.
- 4b. If the email format is invalid, the system requests correction.
- 4c. If the phone number format is invalid, the system requests correction.
- 5a. If the update fails, the system notifies the actor and requests a retry.

2.2.1.2 System Requirements

Based on the defined use cases, system requirements were established to describe both functional capabilities and non-functional constraints. Functional requirements specify what the system must do, while non-functional requirements define quality attributes and operational constraints (ROSENBERG; STEPHENS, 2007).

2.2.1.2.1 Functional Requirements

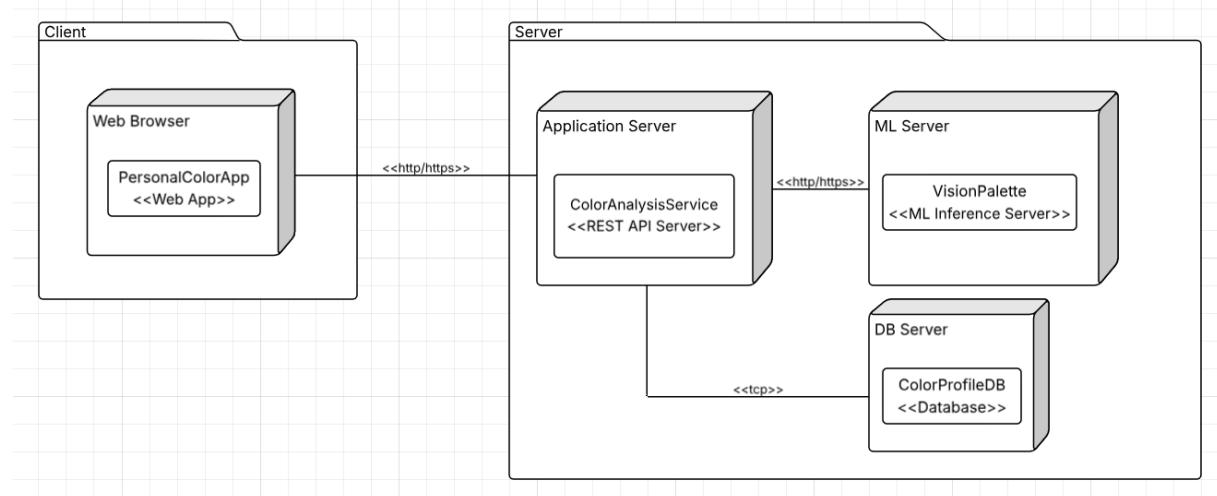
- FR1. The system must allow users to create an account using name, phone number, email address, and password.
- FR2. The system must allow users to register and store payment information.
- FR3. The system must allow users to authenticate using email and password.
- FR4. The system must allow users to upload an image for personal color analysis.
- FR5. The system must allow users to update profile information.
- FR6. The system must provide guidelines for capturing valid images for analysis.
- FR7. The system must analyze uploaded images using a computer vision-based model.
- FR8. The system must display the resulting personal color palette to the user.

2.2.1.2.2 Non-Functional Requirements

- NFR1. The system must be user-friendly and intuitive.
- NFR2. The system must support responsive design across different devices.
- NFR3. The system must ensure data security and user privacy.
- NFR4. The system must provide real-time feedback during the analysis process.
- NFR5. The system must return analysis results within a maximum of 30 seconds.
- NFR6. The system must be scalable to support an increasing number of users.
- NFR7. The system must maintain high availability and reliability.

2.2.2 Architecture and Technology:

Figure 3 - Deployment Diagram



Source: Prepared by the author.

Figure 2 presents the UML deployment diagram of the VisionPalette platform, illustrating the system's runtime components and their physical distribution across processing nodes. Deployment diagrams are used to model the physical architecture of a system, highlighting how software components are allocated to infrastructure resources and how they interact at execution time.

The VisionPalette platform adopts a cloud-based microservices architecture, in which each major system component—frontend, backend, database, and computer vision model service—operates as an independent service. This architectural approach promotes scalability, modularity, and maintainability by establishing clear responsibility boundaries between components. Such separation aligns with established best practices and empirical findings on the advantages of microservices architectures in modern software systems, particularly in terms of fault isolation, independent deployment, and scalability (DRAGONI et al., 2021; TAIBI; LENARDUZZI; PAHL, 2017).

User interaction follows a client–server model, where the Next.js frontend acts as the client and communicates with the NestJS backend, which serves as the central orchestrator of business logic. The backend manages user authentication, payment flow, and data persistence, and forwards image analysis requests to the VisionPalette model service. This separation of concerns ensures that user-facing functionality, application logic, and machine learning inference remain decoupled.

For the minimum viable product (MVP), certain deployment simplifications were intentionally adopted for academic and cost-related reasons. The frontend and backend services were deployed on the same service instance, and the PostgreSQL database was hosted outside the Amazon Web Services (AWS) environment. These decisions reduced infrastructure costs and operational complexity during early development while preserving the logical separation between services. Importantly, these deployment choices do not compromise the architectural principles of modularity or scalability embedded in the system design.

In a production-ready configuration, all system components—including frontend, backend, database, and the VisionPalette model service—are designed to operate as fully independent microservices deployed within the AWS ecosystem. This configuration enables horizontal scalability, independent release cycles, and improved fault tolerance, consistent with best practices in cloud-native system design and microservices-based architectures (BLINOWSKI; OJDOWSKA; PRZYBYŁEK, 2022).

The proposed architecture integrates seamlessly into the partner company's IT ecosystem, as AWS is already established as the organization's standard cloud platform. Deploying the system entirely within AWS enables unified infrastructure management, consistent security policies, secure inter-service communication, and alignment with existing DevOps workflows. These characteristics support operational efficiency, governance, and long-term maintainability, reinforcing the suitability of the chosen architecture for the company's technological environment (THARWANI; PURKAYASTHA, 2024)

2.2.3 Development and Implementation (MVP):

The development of the VisionPalette solution followed an iterative and incremental methodology, aligned with core principles of Agile software development. This approach emphasized continuous validation, stakeholder collaboration, and adaptability to evolving requirements throughout the project lifecycle. Rather than adopting a rigid framework such as Scrum or Kanban, the methodology was applied in a pragmatic manner, consistent with both the academic context and the partner company's operational reality.

Development activities were organized into short cycles, each culminating in validation meetings in which intermediate results were formally presented and discussed with the project partner. These review sessions enabled systematic assessment of progress, early identification of issues, and alignment between technical implementation and business expectations. In addition to scheduled reviews, informal and punctual discussions with the domain specialist occurred throughout the development process to clarify requirements, resolve uncertainties, and address emerging needs.

Feedback collected during validation meetings occasionally led to requirement adjustments or the identification of new features. Proposed changes were evaluated based on their impact on project scope, expected value, and implementation effort. Approved changes were prioritized and incorporated incrementally into subsequent development cycles or considered during planning for the next iteration. This process ensured controlled scope evolution while preserving flexibility.

Continuous improvement was also supported through retrospective discussions conducted at the end of development cycles. During these retrospectives, strengths, limitations, and opportunities for improvement related to tools, workflows, and development practices were identified. The resulting insights were documented and translated into concrete actions, such as workflow refinements or adjustments to development practices. Periodic alignment sessions and project plan reviews further contributed to maintaining quality, efficiency, and coherence with project objectives.

Following the completion and validation of the Convolutional Neural Network (CNN) models, a pilot deployment strategy was defined to integrate the VisionPalette solution into a test environment that closely resembled a production setting. To ensure greater control over computational resources and system stability, the deployment strategy was revised to adopt cloud-based infrastructure using AWS.

The solution was deployed using AWS EC2 instances provisioned through the AWS Learner Lab environment, which provided cloud resources supported by educational credits. This configuration ensured sufficient memory, storage, and processing capacity for both the application layer and the model inference service, while maintaining flexibility for testing and experimentation.

The deployed architecture consisted of two primary services hosted on separate EC2 instances to preserve modularity. The first instance hosted the VisionPalette model inference service, implemented as a RESTful API using FastAPI and containerized with Docker. The second instance hosted the VisionPalette application, encompassing frontend and backend components, also containerized using Docker. This separation enabled independent resource allocation, simplified maintenance, and facilitated future scalability.

Integration between services was achieved through HTTP-based communication, allowing the application layer to transmit image data to the model service and receive classification results in real time. Although GPU acceleration was not available in the pilot environment, CPU-based inference proved sufficient for the project's requirements, achieving response times in the order of hundreds of milliseconds per image and supporting near real-time interaction.

Containerization played a central role in the deployment process by ensuring consistency between development, testing, and deployment environments. Docker was used to package each service along with its dependencies, improving reproducibility, isolation, and operational reliability. Deployment configurations were optimized to include only essential services, reducing resource consumption during the pilot phase.

For data persistence, the system was integrated with a managed PostgreSQL database hosted on Aiven. This managed service provided automated backups, encryption at rest and in transit, monitoring, and high availability, thereby reducing operational overhead and ensuring secure and reliable data storage throughout testing and validation.

Overall, the pilot deployment demonstrated the feasibility of integrating the VisionPalette solution into a cloud-based environment. It validated the system's architectural design, operational performance, and scalability under realistic conditions, providing a solid foundation for future evolution toward a fully production-ready deployment.

2.2.4 Testing and Technical Evaluation:

To evaluate whether the developed solution meets the company's technical requirements and quality standards, a comprehensive testing strategy was adopted. The strategy combined unit testing, end-to-end (E2E) testing, and user-oriented workflow validation, aligning with corporate testing practice, functional validation, and performance verification.

Backend Testing

Unit tests were implemented in the backend using Jest to validate core business logic and application use cases. These tests focused on ensuring correct behavior of individual components in isolation, including user creation, authentication, payment handling, image analysis orchestration, and error handling scenarios. By validating both successful flows and failure cases, backend unit tests enabled early detection of defects and increased confidence in system reliability (CLIFTON, 2017).

Figure 2 presents the results of the backend test suite execution, showing that all defined test cases passed successfully. A total of 10 test suites and 21 individual tests were executed without failures, demonstrating that the backend logic complies with the functional requirements specified in Section 2.2.1, particularly those related to user management, payment processing, and analysis request handling.

Figure 4 – Backend Unit Tests Results

```

PASS | src/modules/user/useCases/getUser/getUser.spec.ts (15.974 s)
Get User
  ✓ Should be able to get user (59 ms)
  ✓ Should be able to throw error if user not exists (21 ms)

PASS | src/modules/auth/useCases/validateUser/validateUser.spec.ts (17.046 s)
Validate User
  ✓ Should be able to return user when credentials are correct (273 ms)
  ✓ Should be able to throw error when credentials are invalid (274 ms)

PASS | src/modules/palette/useCases/getPaletteFromImage/getPaletteFromImage.spec.ts (17.1 s)
Analyze Image
  ✓ Should be able to throw error if VisionPalette fails (47 ms)
  ✓ Should be able to return a color palette (132 ms)
  ✓ Should be able to throw error if palette not exists (310 ms)

PASS | src/modules/payment/useCases/createPaymentMethod/createPaymentMethod.spec.ts
Create Payment Method
  ✓ Should be able to create a payment method (5 ms)
  ✓ Should be able to return an existing payment method when trying to create with same card number (2 ms)

PASS | src/modules/user/useCases/createUser/createUser.spec.ts
Create User
  ✓ Should be able to create user (521 ms)
  ✓ Should be able to create user with password encrypted (748 ms)
  ✓ Should be able to throw error when trying to create user with a registered email (12 ms)

PASS | src/modules/auth/useCases/signIn/signIn.spec.ts
Sign in
  ✓ Should be able to create valid access token (36 ms)

PASS | src/modules/payment/useCases/updatePaymentMethod.ts/updatePaymentMethod.spec.ts
Update Payment Method
  ✓ Should be able to update a payment method (3 ms)
  ✓ Should be able to create a new payment method if not exists (2 ms)
  ✓ Should be able to update a different existing payment method (11 ms)

PASS | src/modules/palette/useCases/createPalettes/createPalettes.spec.ts
Create Palettes
  ✓ Should be able to create palettes (4 ms)

PASS | src/modules/order/useCases/createOrder/createOrder.spec.ts
Create Order
  ✓ Should be able to create order (7 ms)
  ✓ Should be able to throw error if user not exists (14 ms)

PASS | src/modules/user/useCases/updateUser/updateUser.spec.ts
Update User
  ✓ Should be able to update user (11 ms)
  ✓ Should be able to throw error if user not exists (16 ms)

Test Suites: 10 passed, 10 total
Tests:       21 passed, 21 total
Snapshots:   0 total
Time:        23.721 s, estimated 27 s

```

Source: Prepared by the author.

Frontend Testing

Frontend unit tests were developed using Jest and React Testing Library to verify component rendering, form validation, and utility logic. These tests ensured that the user interface behaved as expected under different interaction scenarios, such as invalid inputs, navigation flows, and conditional rendering. Figure 3 illustrates the successful execution of frontend tests, confirming that the interface meets usability and reliability requirements.

Figure 5 – Frontend Unit Tests Results

```

PASS  __tests__/functions.test.tsx
Upload image
  ✓ should accept JPEG, PNG or JPG files (104 ms)
  ✓ Should ignore invalid files (8 ms)
  ✓ Should deal with empty inputs (11 ms)
Verify email format
  ✓ Should return true when format is valid (5 ms)
  ✓ Should return false when format is invalid (3 ms)
Verify password format
  ✓ Should return true when format is valid (2 ms)
  ✓ Should return false when password is less than 8 characteres (1 ms)
  ✓ Should return false when password doesn't have uppercase letter (1 ms)
  ✓ Should return false when password doesn't have special character (3 ms)
  ✓ Should return false when password doesn't have number (1 ms)

PASS  __tests__/render.test.tsx (12.302 s)
Render Components
  ✓ renders the Upload Area (295 ms)
  ✓ renders Header (16 ms)
  ✓ renders Phone Input (78 ms)
  ✓ renders Form Input (58 ms)
  ✓ renders Form Button (47 ms)
Render Home components
  ✓ renders Home Body (43 ms)
  ✓ renders the Menu icon (13 ms)
  ✓ renders the Sidebar (46 ms)
  ✓ not renders the Sidebar when not opened (12 ms)
Render Landing Page components
  ✓ renders Banner (35 ms)
  ✓ renders the Footer (34 ms)
  ✓ renders the Landing page Button (11 ms)

Test Suites: 2 passed, 2 total
Tests:       22 passed, 22 total
Snapshots:   0 total
Time:        20.081 s, estimated 21 s

```

Source: Prepared by the author.

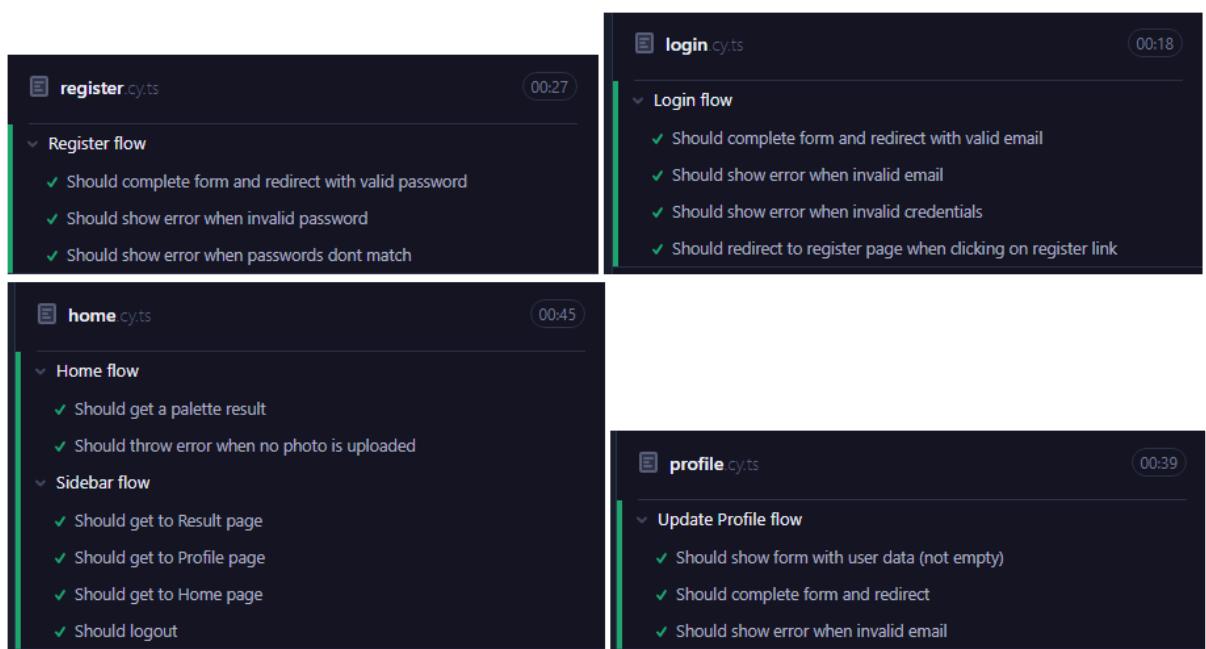
End-to-End Testing and User Acceptance Validation

E2E testing is a software testing approach that validates the complete workflow of an application from start to finish, ensuring that all integrated components—from the user interface to backend services and databases—operate correctly together (THE COMPLETE GUIDE TO END-TO-END TESTING, 2025). The successful execution of these tests indicates that the system is functionally stable and ready for pilot deployment.

E2E tests were implemented using Cypress to simulate real user interactions across the entire application stack. These tests covered complete user journeys, including account registration, login, profile updates, image upload for analysis, and color palette retrieval. By executing these workflows in an integrated environment, E2E testing validated that frontend, backend, database, and external services functioned cohesively.

E2E testing aligns with User Acceptance Testing (UAT) practices, as it verifies that the system behaves as expected from the user's perspective and satisfies business requirements under realistic usage scenarios. Figure 4 presents the results of the E2E test executions, demonstrating that all tested workflows completed successfully without critical errors.

Figure 6 – E2E Tests Results



Source: Prepared by the author.

Performance and Reliability Considerations

Although no dedicated load or stress testing framework was employed during the MVP phase, performance was indirectly validated through system behavior observed during E2E tests and pilot deployment. Image analysis requests consistently returned results within the non-functional requirement of a maximum response time of 30

seconds, with typical inference times in the order of hundreds of milliseconds per image when executed on CPU-based infrastructure.

System reliability was further supported by automated testing coverage and containerized deployment, which ensured consistent execution environments and reduced configuration-related errors.

Evaluation Summary

The testing results demonstrate that the VisionPalette solution meets the company's technical requirements and quality standards. Functional requirements related to user management, payment handling, image analysis, and result presentation were validated through unit and E2E tests. Non-functional requirements concerning usability, responsiveness, reliability, and performance were also satisfied within the scope of the MVP.

Overall, the adopted testing strategy provided sufficient evidence of system correctness, integration stability, and readiness for further evolution toward a production-ready deployment.

2.3 Assessment of Impact and Contribution to the Business

2.3.1 Defining Corporate Success Metrics:

The success of the proposed solution was evaluated using a set of Key Performance Indicators (KPIs) aligned with the objectives defined in Section 1.3. These metrics were selected to assess both the technical effectiveness of the solution and its operational feasibility within the partner company's context.

From a corporate perspective, the primary success criteria focused on the system's ability to significantly reduce the time required to obtain a personal color analysis, while delivering reliable and interpretable results suitable for real-world use. Accordingly, the following KPIs were defined:

- Model Accuracy, measured during training, testing, and validation phases, to evaluate the system's ability to correctly classify personal color palettes.

- Validation Accuracy on Real-World Images, assessed separately for challenging images with noise and heterogeneous lighting conditions and for images with clearer patterns and controlled lighting, to evaluate robustness.
- Top-2 Accuracy, measuring whether the correct class appears among the two highest-probability predictions. This metric was particularly relevant given the inherent visual similarity between certain color palettes and the exploratory nature of the solution.
- Inference and Response Time, measuring the time required by the model API to process an image and return a result, to assess compliance with operational and usability requirements.

These KPIs enabled a structured comparison between the baseline scenario, characterized by a fully manual and consultant-dependent analysis process, and the post-implementation scenario, in which automated classification is used to generate results within minutes. While the manual process does not provide quantitative accuracy metrics, it establishes a baseline in terms of time, cost, and scalability, against which the automated solution can be evaluated.

Measurement Methodology

To measure the defined KPIs, a combination of controlled datasets and real-world validation data was employed. The training and testing datasets were composed of artificially generated images, created using generative artificial intelligence tools such as ChatGPT and Gemini. These images followed strict visual constraints—including portrait framing, neutral white background, and gray clothing—to reduce variability and allow the model to focus on facial and color-related features.

In total, 510 images were used during the training and testing phases, with 306 images allocated for training and 204 images reserved for testing. This split ensured sufficient data for model learning while preserving an independent test set for performance evaluation.

To assess generalization and robustness, two additional validation scenarios were defined using real-world images obtained from professional consultancy sources and publicly available datasets. The first validation set consisted of 48 images with high variability, including inconsistent lighting, image noise, and class imbalance. These

characteristics were intentionally preserved to simulate adverse real-world conditions. A second validation scenario included images with clearer facial patterns, consistent lighting, and higher visual quality, enabling comparative analysis under more favorable conditions.

For each dataset split, top-1 accuracy and Top-2 accuracy were computed independently to evaluate learning stability and classification ambiguity. This separation allowed the identification of performance degradation caused by noise, lighting variation, and class imbalance.

Inference and response time measurements were obtained by monitoring the execution time of the model API during prediction requests. Measurements were collected across five runs to ensure consistency and to evaluate the system's suitability for near real-time interaction. These results were compared against the non-functional requirement defined in Section 2.2.1, which established a maximum acceptable response time of 30 seconds.

By combining synthetic training data, real-world validation scenarios, and multiple performance indicators, the adopted measurement methodology enabled a comprehensive assessment of the solution's effectiveness, limitations, and practical applicability. This approach supports an objective evaluation of the system's contribution relative to the baseline manual process and provides a solid foundation for interpreting the results presented in the following sections.

2.3.2 Results and Impact Analysis:

This section presents the results of the solution implementation and analyzes its impact relative to the baseline metrics defined in Section 1.2. The baseline scenario is characterized by a fully manual personal color analysis process, with an average duration of approximately 120 minutes per session, high dependency on professional expertise, and limited scalability. The proposed solution introduces an automated, AI-based classification component, enabling near real-time inference and supporting significant operational changes.

Table 1 – Accuracy Results for Each Approach

Model	Test Accuracy	Validation Accuracy (set 1)	Validation Accuracy (set 2)	Top-2
V1	55.9%	4.17%	—	8.33%
Gating	55.9%	4.17%	—	4.17%
4 Seasons	71.1%	41.67%	50%	62.5% ~ 66.67%
Multi-Models	87.3% / 75.5%*	43.75%	66.67%	—

*Temperature model / Intensity model

Source: Prepared by the author.

Table 1 summarizes the comparative performance of the evaluated modeling strategies across test and validation datasets. Several key observations emerge from this comparison.

The 12-class prediction model and the gating-based architecture exhibited limited robustness when applied to real-world data. Although both approaches achieved moderate accuracy on the controlled test set, their performance collapsed on validation datasets, indicating high sensitivity to noise, lighting variability, and subtle visual differences between classes. These results suggest that direct prediction of twelve fine-grained palettes is not well suited to the current dataset size and variability.

The 4-class model, which predicts only the four main seasons, demonstrated a substantial improvement in generalization performance. By reducing classification granularity, this approach achieved significantly higher validation accuracy on both real-world datasets. This result indicates that coarse-grained seasonal classification is more robust under realistic conditions and better aligned with the visual separability present in the data.

The multi-model approach achieved the strongest overall results. By decomposing the task into two interpretable binary classifications—temperature (warm vs. cool) and intensity (light vs. deep)—the system was able to capture key perceptual attributes more reliably. When combined, these predictions produced the highest validation accuracy across all evaluated strategies, particularly on the higher-quality real-world dataset. This confirms that problem decomposition can mitigate ambiguity and improve robustness in visually subtle classification tasks.

The analysis of Top-2 accuracy provides additional insight into model behavior. Even in cases where the top-1 prediction was incorrect, the correct class frequently appeared among the two most probable outputs. This characteristic is especially relevant in personal color analysis, where boundaries between palettes are inherently subtle and overlapping. From an application perspective, this supports the feasibility of using model output as a guided decision aid rather than a rigid classification.

In terms of operational performance, inference and response time measurements demonstrate that the solution comfortably meets the non-functional requirements defined in Section 2.2.1. As shown in Table 2, total API response times remained below 1 second on average, even without GPU acceleration. This represents a dramatic improvement relative to the baseline manual process and confirms the technical feasibility of near real-time color palette generation.

Table 2 – Inference and Response Time for Each Model

Model	Inference Time (ms)	Response Time (ms)
4 Seasons	143	823
Multi-Models	281.2	976.024

Source: Prepared by the author.

When compared to the baseline manual process—requiring approximately 120 minutes per session—the automated inference component reduces the core analytical step to sub-second execution time. Although the full service design may still include optional human interpretation or guidance, the automation of the classification step alone represents a reduction of well over 75% in the time required to obtain an initial result. This satisfies and exceeds the primary quantitative objective established in Section 1.3.

Additionally, the low inference latency and scalable architecture enable a cost structure that is independent of consultant availability, supporting higher throughput without proportional increases in operational cost.

Beyond quantitative performance, the results indicate several qualitative gains for both the organization and end users. From a business perspective, the solution

increases agility by enabling rapid initial assessments, reduces reliance on highly specialized human resources for preliminary analysis, and supports consistent decision-making across sessions.

From a user perspective, the availability of fast, automated feedback enhances accessibility and autonomy, allowing individuals to explore their personal color palette without the time and logistical constraints of traditional consultations. The interpretability of the multi-model approach, combined with Top-2 predictions, also supports transparency and user trust, as results can be understood as probabilistic guidance rather than absolute judgments.

Overall, the results demonstrate that the proposed solution significantly mitigates the core limitations of the baseline manual process, particularly in terms of efficiency, scalability, and operational feasibility. However, the findings also indicate that, given the current accuracy levels and sensitivity to real-world variability, the solution does not yet fully support a completely autonomous, user-ready replacement for professional personal color analysis.

Instead, the system proves to be most effective as an auxiliary decision-support tool, capable of accelerating the initial stages of analysis, reducing manual workload, and providing consistent, interpretable guidance. These results validate the technical feasibility of applying artificial intelligence to personal color classification while clearly delineating the constraints that must be addressed—such as dataset expansion, robustness to noise, and improved generalization—before full automation can be reliably achieved.

2.3.3 Cost-Benefit Analysis:

This section presents an estimated cost–benefit analysis of deploying the proposed solution in a production environment. The analysis considers infrastructure costs, operational assumptions, and a simplified revenue model aligned with the exploratory and MVP-oriented nature of the project.

The estimated production infrastructure costs were calculated using the AWS Pricing Calculator, considering on-demand pricing and continuous operation. The proposed architecture includes separate instances for application services and model

inference, as well as a managed relational database. The estimated monthly costs are summarized below.

- Application Layer (Frontend + Backend)
 - Two EC2 instances of type *t3.medium* were considered, one hosting the frontend and the other hosting the backend services.
 - 2 instances × USD 0.0416/hour × 730 hours/month
 - Estimated monthly cost: USD 60.74
- Model Inference Service
 - One EC2 instance of type *t3.large* was allocated for the VisionPalette model service.
 - 1 instance × USD 0.0832/hour × 730 hours/month
 - Compute cost: USD 60.74
 - EBS storage: 24 GB × USD 0.08/GB/month = USD 1.92
 - Total model service cost: USD 62.66/month
- Database Layer (RDS)
 - One managed PostgreSQL instance was considered.
 - Compute: USD 13.14/month
 - Storage: 20 GB × USD 0.115/GB/month = USD 2.30
 - Total database cost: USD 15.44/month

Based on these estimates, the total monthly infrastructure cost is approximately USD 138.84, corresponding to an annual cost of approximately USD 1,666.08.

No additional licensing costs were incurred, as all major technologies used in the project are open source. Development labor costs were not included in this calculation, as the project was conducted in an academic context and represents sunk development effort rather than an operational expense.

To evaluate financial feasibility, a simplified revenue model was considered based on a pay-per-analysis pricing strategy, aligned with the system design (one image per payment). A conservative fee of R\$ 20.00 per image was assumed. This value is significantly lower than the typical cost of a full professional personal color consultation, which generally ranges between R\$ 450.00 and R\$ 850.00, positioning the tool as a highly accessible alternative or complementary service.

Assuming an exchange rate of approximately R\$ 6.00 per USD, the estimated monthly infrastructure cost of USD 138.84 corresponds to approximately R\$ 833.04 per month.

Under this pricing model, the system would need to process approximately:

- 42 analyses per month ($R\$ 833.04 \div R\$ 20.00$)

to fully cover infrastructure costs and reach operational break-even.

This volume corresponds to roughly one to two analyses per day, which remains significantly below the throughput limitations of traditional manual personal color consultations. Even modest user adoption beyond this threshold would generate positive operational margins, particularly given the low marginal cost per additional analysis.

2.3.4 Critical Success Factors and Lessons Learned:

The implementation and evaluation of the proposed solution revealed a set of critical success factors and limitations that directly influenced its impact on the partner company's results. These factors played a central role in shaping the evolution of the project from its initial objective toward a more realistic and applicable outcome.

One of the most significant success factors was the structured modeling of the personal color analysis domain. By aligning machine learning strategies with well-established analytical dimensions—such as temperature, intensity, depth, and contrast—the project ensured conceptual coherence between professional consultancy practices and computational modeling. This alignment enabled the exploration of alternative architectures, particularly the multi-model approach, which proved more robust and interpretable than monolithic classification strategies.

Another key success factor was the decision to validate the solution using real-world images, including challenging scenarios characterized by noise, inconsistent lighting, and heterogeneous visual patterns. Although this choice exposed limitations in model generalization, it provided an accurate assessment of real deployment conditions and prevented overly optimistic conclusions based solely on controlled test data. This validation strategy contributed to a realistic understanding of the solution's capabilities and constraints.

From an operational perspective, architectural and infrastructure decisions also contributed positively to the project's impact. The cloud-based, modular architecture enabled low-latency inference and demonstrated that significant reductions in analysis time—exceeding the initially defined 75% target—are technically feasible. These results validate the system's potential to support scalable and cost-efficient workflows, even in the absence of full automation.

At the same time, several lessons learned emerged during implementation. Dataset representativeness proved to be the primary limiting factor affecting model performance. The pronounced discrepancy between test accuracy on controlled datasets and performance on real-world validation sets highlights the need for datasets that more accurately reflect the diversity of skin tones, lighting conditions, image quality, and contextual variability present in real consultations. This limitation directly impacted the feasibility of achieving the initial goal of fully automatic, user-ready personal color classification.

Additionally, the results demonstrated that high-granularity classification into twelve seasonal palettes is particularly sensitive to visual ambiguity and data scarcity. Even with advanced architectures, fine-grained distinctions between palettes remain challenging due to overlapping visual characteristics. This insight motivated a shift in perspective, from attempting to fully replace professional analysis to positioning the solution as an auxiliary tool that supports and accelerates decision-making.

As a result of these findings, the project's practical contribution evolved from its original objective. Rather than functioning as a fully autonomous replacement for personal color consultancy, the solution proved more effective as a decision-support system, capable of delivering fast, preliminary classifications that narrow the analytical space and reduce manual workload. In this role, the system can significantly shorten the initial stages of analysis, support consultant efficiency, and improve scalability, while preserving the complementary value elements of professional consultancy.

Overall, the critical success factors and lessons learned demonstrate that the project achieved its exploratory and applied research goals. By identifying both the technical feasibility and the boundaries of automated personal color analysis, the work

provides a solid foundation for future improvements and informed integration of artificial intelligence into the partner company's service offerings.

3 Conclusion

This project set out to investigate the feasibility of developing an automated personal color classification system capable of operating as a fully independent, user-ready application, without the need for professional consultancy guidance. Based on the results obtained, this primary objective was not fully achieved. The evaluation demonstrated that, given the current levels of accuracy and sensitivity to real-world variability, the solution does not yet provide sufficient reliability to replace professional personal color analysis in a completely autonomous manner.

Despite this limitation, the project achieved significant and relevant results. The developed system successfully demonstrated that artificial intelligence and computer vision techniques can be effectively applied to personal color analysis as an auxiliary decision-support tool. By automating the initial analytical stage, the solution is capable of delivering preliminary classifications in near real time, substantially reducing the effort required in the early phases of the process. Even without full automation, the solution enables a reduction in analysis time well beyond the initially defined target of 75%, validating its operational value.

From a business perspective, the project delivers meaningful impact. The proposed solution introduces a scalable and low-cost digital service that can expand access to personal color analysis, attract new users, and support higher-value services offered by the company. By reducing dependence on fully manual processes, the system allows consultants to focus on interpretative, strategic, and creative aspects of image consultancy, such as personalized guidance, styling, and content creation. This hybrid model strengthens operational efficiency while preserving the differentiated value of professional consultancy.

Throughout the development and evaluation process, the project highlighted the critical importance of dataset quality and representativeness. The disparity observed between controlled test performance and real-world validation results reinforces that data diversity—encompassing lighting conditions, skin tones, image quality, and environmental variability—is a determining factor for the success of machine learning

systems in this domain. These findings represent a central contribution of this work, as they clarify the practical limitations of high-granularity classification under constrained data conditions.

Based on these insights, several recommendations can be made for the future evolution of the solution. From a technical standpoint, future work should prioritize the expansion and curation of real-world datasets, incorporating a broader range of demographic and environmental conditions. Additional improvements may include the adoption of data augmentation strategies, domain adaptation techniques, confidence-based output thresholds, and explainability mechanisms to improve user trust. From a product perspective, the system could evolve toward hybrid workflows that combine automated pre-classification with optional professional validation, aligning technological efficiency with service quality.

Regarding scalability and maintenance, the cloud-based and modular architecture implemented in this project provides a solid foundation for future growth. The separation between application services and model inference enables independent updates, retraining, and deployment cycles. As model performance improves, the same infrastructure can support increased usage with minimal additional cost, reinforcing long-term sustainability.

Finally, knowledge transfer was an integral aspect of this project. Technical documentation covering system architecture, deployment configuration, model interfaces, and data pipelines was produced to support internal understanding and future development by the company's technical team. This documentation ensures continuity, facilitates maintenance, and enables further experimentation and refinement of the solution beyond the scope of this academic work.

In summary, while the original goal of full automation was not completely achieved, the project successfully demonstrated the feasibility, value, and limitations of applying artificial intelligence to personal color analysis. By delivering a functional MVP, identifying critical constraints, and establishing a clear roadmap for evolution, this work provides a robust foundation for future research and practical innovation in the intersection of fashion, image consultancy, and artificial intelligence.

References

- BROWN, J. Wendy; ROJAS, Andrea. **Determining Personal Colors Guide C-315.** [s.l.: s.n.]. Available at: https://pubs.nmsu.edu/_c/C315.pdf. Accessed on Apr. 29, 2025.
- XU, Pingyuan *et al.* Applications of artificial intelligence and machine learning in image processing. **Frontiers in Materials**, v. 11, p. 1431179, 2024.
- Building Your Application: Rendering | Next.js.** Available at: <https://nextjs.org/docs/app/building-your-application/rendering>. Accessed on Apr. 29, 2025.
- BEIGHTON, B. **Why I choose NestJS for a SaaS API.** Available at: <https://bradbeighton.medium.com/nestjs-the-pros-and-cons-aff714607b07>. Accessed on Apr. 29, 2025.
- Benchmarks - FastAPI.** Available at: <https://fastapi.tiangolo.com/benchmarks/#benchmarks-and-speed>. Accessed on Apr. 29, 2025.
- SIMONYAN, Karen; ZISSERMAN, Andrew. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.
- WU, Songtao; ZHONG, Shenghua; LIU, Yan. Deep residual learning for image recognition. **Multimed. Tools Appl**, p. 1-17, 2017.
- TAN, Mingxing; LE, Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In: **International conference on machine learning**. PMLR, 2019. p. 6105-6114.
- WEISS, Karl; KHOSHGOFTAAR, Taghi M.; WANG, DingDing. A survey of transfer learning. **Journal of Big data**, v. 3, n. 1, p. 9, 2016.
- ROSENBERG, Doug; STEPHENS, Matt. **Use case driven object modeling with UML: theory and practice.** Berkeley, CA: Apress, 2007.
- M. Mazzara, N. Dragoni, A. Buccharone, A. Giaretta, S. T. Larsen and S. Dustdar, "Microservices: Migration of a Mission Critical System," in **IEEE Transactions on Services Computing**, vol. 14, no. 5, pp. 1464-1477, 1 Sept.-Oct. 2021, doi: 10.1109/TSC.2018.2889087.
- TAIBI, Davide; LENARDUZZI, Valentina; PAHL, Claus. Processes, motivations, and issues for migrating to microservices architectures: An empirical investigation. **IEEE Cloud Computing**, v. 4, n. 5, p. 22-32, 2017.
- BLINOWSKI, Grzegorz; OJDOWSKA, Anna; PRZYBYŁEK, Adam. Monolithic vs. microservice architecture: A performance and scalability evaluation. **IEEE access**, v. 10, p. 20357-20374, 2022.

THARWANI, Jay; PURKAYASTHA, Arnab A. Cost-performance evaluation of general compute instances: Aws, azure, gcp, and oci. **arXiv preprint arXiv:2412.03037**, 2024.

CLIFTON, M. **Unit Testing Succinctly**. [s.l]: s.n.]. p. 1-128

The Complete Guide to End-to-End Testing. [s.l]: s.n.]. Available at: https://spring2019.stpcon.com/wp-content/uploads/2018/12/SB_EBK_End-to-End-Testing.pdf . Accessed on June 26, 2025.