

Faculdade de Engenharia da Universidade do Porto



FEUP

**Utilização de conceitos de Ambient Intelligence
em aplicação NFC**

Tiago André Oliveira Exposto

Mestrado Integrado em Engenharia Electrotécnica e de
Computadores

Junho 2011

Utilização de conceitos de Ambient Intelligence em aplicação NFC

Tiago André Oliveira Exposto

Orientador

Mário Sousa

Orientador CreativeSystems

João Sousa

Resumo

Hoje em dia a tecnologia está inteiramente ligada às actividades da sociedade no dia-a-dia e o humano interage em ambientes computacionais que são inteligentes e adaptáveis ao sistema. Este tipo de ambiente é apelidado de Ambiente *Inteligente (Aml)*. O Ambiente *Inteligente* promove ambientes em que os humanos encontram-se rodeados de interfaces inteligentes, suportadas por computação e tecnologias de rede que estão incorporadas nos objectos do quotidiano. Estes ambientes são equipados com dispositivos electrónicos e serviços não-intrusivos que são sensíveis às necessidades das pessoas, personalizados aos seus requerimentos, sensíveis à sua presença, e antecipatórios ao seu comportamento.

Os dispositivos móveis, maioritariamente telemóveis, são os maiores impulsionadores do Ambiente *Inteligente* já que são dispositivos com múltiplas funções e para além disso estão presentes ou por perto da maior parte da população. À medida que os telemóveis se vão tornando cada vez mais indispensáveis em todo o mundo, o uso de *NFC (Near-Field Communication)* está a emergir como uma solução prometedora para o Aml. O *NFC* é uma tecnologia de comunicação de curto-alcance que permite a troca de informações entre dois dispositivos.

É neste contexto que se engloba este projecto, no qual se pretende criar uma aplicação inteligente para um dispositivo móvel com tecnologia *NFC* que permita ao utilizador obter sugestões dependendo do item seleccionado. Assim sendo quando o dispositivo móvel é aproximado dum posto de recepção *NFC*/item com etiqueta *NFC* a aplicação obtém informações por via *GSM/GPRS/3G/Wi-Fi*, de dados relativos ao item e outras sugestões relevantes para o utilizador. Inicialmente cada utilizador terá de fazer um registo com o nome, idade, profissão, entre outros. A partir daí é criado um perfil para o cliente que de acordo com a sua actividade vão sendo actualizadas as sugestões, utilizando algoritmos de clustering. Os algoritmos devem ser capazes de perceber o ambiente e, a partir da interacção do utilizador com o ambiente, adquirir implicitamente percepções relativas ao examinado e assim antecipar o utilizador ao sugerir possíveis opções.

Abstract

Nowadays technology is entirely connected to the activities of the society in day-by-day and the human being interacts in computational environments that are intelligent and adaptable to the system. This type of environment is called Ambient Intelligence (Aml). Ambient Intelligence promotes environments where the human beings are encircled in intelligent interfaces, supported by computation and network technologies that are incorporated in the objects of the everyday. These environments are equipped with electronics devices and non intrusive services that are sensible to the needs of the people, personalized to them requirements, sensible to them presence, and anticipatory to their behavior.

The mobile devices, mostly cell phones, are the biggest boosters of Ambient Intelligence since they are devices with multiple functions and in addition they are nearby to most of the population. Over time cell phones are becoming more indispensable in the whole world, and the NFC use (Near-Field Communication) is emerging as a promising solution for the Aml. NFC is a short-range communication technology that allows the exchange of information between two devices.

It is this context that the project covers, in which it intends to create an intelligent application for a mobile device with NFC technology that allows the user to get suggestions depending on the item selected. Thus being when the mobile device is approached to a item with a NFC tag the application gets information, by GSM/GPRS/3G/Wi-Fi, of the item selected and gives other pertinent suggestions to the user. Initially each user will have to make a registration with the name, age, job, among others. From then on a profile for the customer is created whom, in accordance with its activity, it updates the suggestion being made, using for that clustering algorithms. The algorithms must be capable to perceive the environment and, from the user interaction with the environment, to implicitly acquire relevant perceptions of the user and as a result anticipate the user behavior suggesting possible options.

Agradecimentos

Esta dissertação é o trabalho final do mestrado integrado, e também o fim de um ciclo da vida académica. Gostaria por isso de agradecer não só as pessoas que me ajudaram na dissertação mas também as que me acompanharam ao longo destes 5 anos.

Quero agradecer ao meu orientador, Mário Sousa, e ao orientador da CreativeSystems, João Sousa, pela orientação e disponibilidade demonstrada.

Também quero agradecer ao Leandro Soares e ao José Mota pelo apoio prestado no desenvolvimento da aplicação.

Agradeço ao professor Rosaldo Rossetti e à professora Ana Paula por me auxiliarem e apoiarem, principalmente na parte dos algoritmos de aprendizagem.

Agradeço à minha família que me deu sempre o apoio e confiança necessário ao longo de todo o percurso académico.

Finalmente gostaria de agradecer aos meus amigos pelo companheirismo e amizade ao longo destes anos.

Índice

Índice de Figuras	XI
Índice de Tabelas.....	XIII
Abreviaturas e Símbolos.....	XV
1. Introdução	1
1.1. Motivação.....	1
1.2. Objectivos.....	1
1.3. Estrutura da tese	2
2. Revisão do Estado de Arte	3
2.1. Comunicações sem fios baseadas em radiofrequências.....	3
2.1.1. NFC	3
2.1.2. RFID	6
2.1.3. Comparação entre o NFC e o RFID	7
2.2. Data mining	9
2.2.1. Clustering	9
2.2.2. K-means clustering	10
2.2.3. Cobweb.....	10
2.2.4. DBScan.....	11
2.2.5. EM Clustering	11
3. Especificação e Definição do Sistema	12
3.1. Levantamento de Requisitos.....	12
3.1.1. Casos de uso	12
3.2. Arquitectura do sistema.....	14
3.2.1. Componente Cliente	15
3.2.2. Componente Servidor	15
4. Desenvolvimento	16
4.1. Hardware utilizado	16
4.2. Ferramentas/linguagens de programação	18
4.2.1. Eclipse IDE + Android SDK	18
4.2.2. Xampp	18
4.2.3. WEKA.....	18
4.2.4. Java	18
4.2.5. SQL.....	19
4.3. Base de dados	20
4.3.1. Modelo de dados.....	20
4.3.2. Dados iniciais.....	21
4.4. Interface do telemóvel	23

4.5.	Protocolo de comunicação.....	25
4.6.	Algoritmo.....	29
5.	Resultados/testes.....	30
5.1.	Clustering	30
5.2.	Validação dos casos de uso	38
5.2.1.	Gestão do Login.....	38
5.2.2.	Efectuar Logout	38
5.2.3.	Gestão do Registo	38
5.2.4.	Leitura da Etiqueta NFC.....	39
5.2.5.	Gestão dos dados dos produtos.....	39
5.2.6.	Efectuar Compra.....	39
5.2.7.	Gerir Endereço de Comunicação.....	39
6.	Conclusões.....	41
6.1.	Conclusão	41
6.2.	Implementações Futuras.....	42
7.	Bibliografia	43
	ANEXO I – TABELAS BASE DE DADOS	44

Índice de Figuras

Imagem 2-1 - a) Telemóvel - Etiqueta b) Leitor - Etiqueta.....	4
Imagem 2-2 - Modos de Operação NFC (8).....	5
Imagem 2-3 - Etiquetas RFID (9).....	6
Imagem 3-1 - Casos de uso	12
Imagem 3-2 - Arquitectura do sistema	14
Imagem 4-1 - Google Nexus S	17
Imagem 4-2 - Modelo de dados	20
Imagem 4-3 - Menu Login	23
Imagem 4-4 - Menu Registo	23
Imagem 4-5 - Menu Em Espera	24
Imagem 4-6 - Menu Ver Produto	24
Imagem 4-7 - Menu Opções.....	24
Imagem 4-8 - Exemplo operação Login.....	25
Imagem 4-9 - Exemplo operação Registo.....	26
Imagem 4-10 - Exemplo operação Buscar Produto.....	26
Imagem 4-11 - Exemplo operação Sugestões	27
Imagem 4-12 - Exemplo operação Compra.....	27
Imagem 5-1 – Clustering DBScan de 6 pontos	31
Imagem 5-2 – Clustering DBScan de 2 pontos	31
Imagem 5-3 - Clustering Cobweb	32
Imagem 5-4 - Clustering Hierarchical de 5 clusters.....	32
Imagem 5-5 - Clustering k-means de 5 clusters	32
Imagem 5-6 - Clustering EM de 5 clusters	33
Imagem 5-7 - Clustering K-means com 6 clusters	36
Imagem 5-8 - Clustering K-means com 7 clusters	36
Imagem 5-9 - Clustering K-means com 8 clusters.....	37

Índice de Figuras

Índice de Tabelas

Tabela 2-1- Diferenças entre etiquetas activas e passivas (12)	7
Tabela 2-2 - Tabela comparativa NFC vs RFID (13)	8
Tabela 4-1 - Preferências do histórico.....	22
Tabela 5-1- Preferências do utilizador 1	33
Tabela 5-2 - Sugestões obtidas para utilizador 1 com 5 clusters	34
Tabela 5-3 - Sugestões obtidas para utilizador 1 com 6 clusters	34
Tabela 5-4 - Sugestões obtidas para utilizador 1 com 7 clusters	34
Tabela 5-5 - Preferências do utilizador 2	34
Tabela 5-6 - Sugestões obtidas para utilizador 2 com 5 clusters	35
Tabela 5-7 - Sugestões obtidas para utilizador 1 com 6 clusters	35
Tabela 5-8 - Sugestões obtidas para utilizador 1 com 7 clusters	35

Índice de Tabelas

Abreviaturas e Símbolos

Aml	Ambient Intelligence
FeliCa	Felicity Card
GNU	General Public License
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
IDE	Integrated Development Kit
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
NFC	Near-Field Communication
RFID	Radio Frequency Identification
SDK	Software Development Kit
SIP	Session Initiation Protocol
SQL	Structured Query Language
UHF	Ultra High Frequency
VOIP	Voice Over IP
WEKA	Waikato Environment for Knowledge Analysis

Abreviaturas e Símbolos

1. Introdução

Este capítulo tem como objectivo introduzir o propósito deste trabalho, incluindo a motivação e os objectivos, assim como uma breve descrição de cada capítulo incluído.

1.1. Motivação

Actualmente a tendência é de se integrar sistemas computacionais ao nosso dia-a-dia, visando a adaptação de tecnologias inovativas para criar ambientes que auxiliam o utilizador nas suas actividades. O conceito de Ambient Intelligence diz exactamente isso, e mecanismos de inferência são fundamentais para dar resultados que possam ser significativos para o utilizador. Para serem feitas as inferências é preciso avaliar os dados dos utilizadores sendo um dos maiores desafios tornar esses dados em conhecimentos e informação relevante para o contexto inserido. A integração disto com o Near Field Communication, tecnologia de comunicação de curto-alcance sem fios que está actualmente a ser comercializada com mais intensidade no mercado móvel, torna este projecto desafiante e permite avaliar as capacidades obtidas durante o curso.

A empresa CreativeSystems já tinha grandes conhecimentos de NFC e portanto o seu principal objectivo era obter conhecimentos sobre Ambiente Intelligence passando por isso a grande motivação desta dissertação por desenvolver uma aplicação para telemóvel que permita ao utilizador obter sugestões inteligentes tendo por base o seu perfil. As sugestões vão ser dadas tendo em conta os históricos de compra dos utilizadores e os seus respectivos perfis, assim como vai depender do produto por ele seleccionado.

1.2. Objectivos

O objectivo desta dissertação passa pelo desenvolvimento duma aplicação para telemóvel com tecnologia NFC integrada que permita o utilizador obter sugestões inteligentes de acordo com o produto pesquisado e com o seu perfil. Para isso vai ter de ser criado um base de dados de utilizadores fictícia com respectivo históricos de compras de modo a ser possível comparar com os dados do utilizador e retornar sugestões relevantes para o utilizador. Para além disso vai ter de ser desenvolvida uma aplicação servidor aparte do telemóvel onde irá estar implementado o algoritmo a partir do qual são feitas as sugestões.

1.3. Estrutura da tese

A dissertação está dividida da seguinte maneira:

- **Capítulo 1** – Introdução do tema da dissertação, a motivação para ela e os princípios objectivos que espero atingir.
- **Capítulo 2** – Estado de Arte, faz uma exposição da tecnologia NFC e tecnologias concorrentes. Além disso é feita uma apresentação do data mining e respectivas técnicas de clustering.
- **Capítulo 3** – Especificação e Definição do Sistema, é elaborado um levantamento de requisitos, é apresentado um diagrama da arquitectura do projecto e é feita uma descrição dos componentes cliente e servidor.
- **Capítulo 4** – Desenvolvimento, inicialmente é feita uma exposição das linguagens de programação e das ferramentas computacionais que vão ser utilizadas e também é definido o hardware sobre a qual a aplicação foi desenvolvida. De seguida é descrito todas as etapas de implementação do projecto formado pelo modelo de dados, a interface do telemóvel, o algoritmo de clustering e protocolos de comunicação.
- **Capítulo 5** – Resultados/testes, é feita a validação dos casos de uso. Também são feitos testes com vários tipos de clustering e é definido qual o melhor e explicado a razão.
- **Capítulo 6** – Conclusões, capítulo final em que é feito um sumário do projecto desenvolvido, e é feita uma descrição crítica dos resultados obtidos. De seguida é referido as implementações a desenvolver no futuro.

2. Revisão do Estado de Arte

Neste capítulo irão ser abordadas mais aprofundadamente as tecnologias que vão ser utilizadas para este projecto. Assim sendo irão ser analisados o Near-field communication e o Data Mining. Para além disso essas tecnologias irão ser confrontadas com outras soluções possíveis que servem o mesmo propósito, tal como o RFID (Radio Frequency Identifier).

2.1. Comunicações sem fios baseadas em radiofrequências

2.1.1. NFC

O NFC ou Near-Field Communication é uma tecnologia para a comunicação sem fios de curto-alcance (10 centímetros). O NFC é uma tecnologia que está em constante expansão e é uma combinação de tecnologias de intercomunicação com o RFID (Radio Frequency Identification). Esta tecnologia permite uma interacção bidireccional simples e intuitiva entre dispositivos electrónicos e possibilita a conexão entre dispositivos com um simples toque ou passagem (1). O número de aplicações para o NFC tem aumentado, sendo especialmente usadas em conjunto com telemóveis. As áreas onde existem mais aplicações são: (2)

- **Pagamentos**
O NFC permite ao utilizador fazer compras rápida e eficazmente, possibilita o acto de compra com dinheiro electrónico, e também a compra de bilhetes electrónicos, como por exemplo para concertos, ou para viagens.
- **Chaves electrónicas**
Por exemplo, chaves do carro, casa ou escritório.
- **Identificação**
O NFC permite ao utilizador usar o seu telemóvel como documento de identificação, como por exemplo como cartão de estudante.
- **Receber e enviar informação**
Permite aceder a dados armazenados em objectos com etiquetas (tais como, posters ou cd's de musica) e assim fazer download de trailers de filmes, mapas de cidade ou extractos musicais.

- **Serviço de setup**

Para evitar configurações complicadas, o NFC pode ser usado para fazer a ligação entre outras tecnologias sem fios de longo-alcance, tais como o Bluetooth ou LAN sem fios.

O NFC é uma tecnologia open-source desenvolvida em conjunto pela Philips e Sony em 2002. O NFC actua na banda de frequência dos 13,56 MHz, e tem uma capacidade de transferência de dados até 424 Kilobits por segundo (3). Foi deliberadamente desenvolvida para ser compatível com etiquetas RFID que operam na banda acima referida (ISO 14443), mas é incompatível com o standard global da EPC. (4)

Na tecnologia NFC é comunicação é sempre feita entre dois dispositivos: (5)

- **O Initiator**

Inicia e controla a informação que vai ser trocada.

- **O Target**

Responde ao requerimento do Initiator.

Por outras palavras, a etiqueta NFC só pode funcionar como Target e todos os outros dispositivos móveis podem funcionar como Initiator ou Target (ver Imagem 2-1).

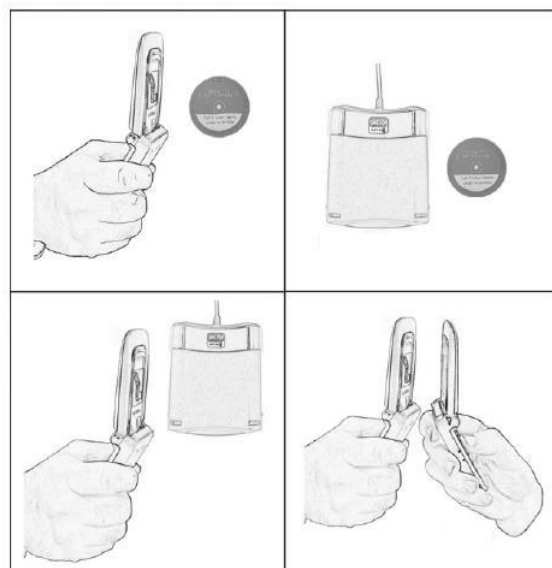


Imagem 2-1 - a) Telemóvel - Etiqueta b) Leitor - Etiqueta
c) Telemóvel - Leitor d) Telemóvel – Telemóvel (6)

Os dispositivos NFC funcionam em três modos de operação e são os únicos dispositivos que permitem a mudança de modos no mesmo dispositivo. Os diferentes modos de operação baseiam-se na ISO/IEC 18092 NFC IP-1 e na ISO/IEC 1443, que são standards dos smartcards contactless. Os três modos de operação são os seguintes (Imagem 2-2): (7)

- Modo leitura/escrita**
 O dispositivo , que é o Initiator, é capaz de ler etiquetas NFC, como por exemplo num poster. Este modo está em conformidade com o ISO 14443 e o FeliCa.
- Modo peer-to-peer**
 Dois dispositivos podem trocar dados, por exemplo podem compartilhar parâmetros Bluetooth ou WiFi ou trocar dados, tais como cartões virtuais ou fotos digitais. Neste caso os dois dispositivos são Initiator e Target dependendo se estão a pedir informação ou a responder a um pedido respectivamente.
- Modo Emulação de Cartão**
 O dispositivo NFC funciona como uma etiqueta NFC, ou seja funciona como Target, tal como nos tradicionais contactless smart card. Este modo é usado principalmente quando não há mudanças na infra-estrutura já existente, como por exemplo nos pagamentos ou compra de bilhetes.

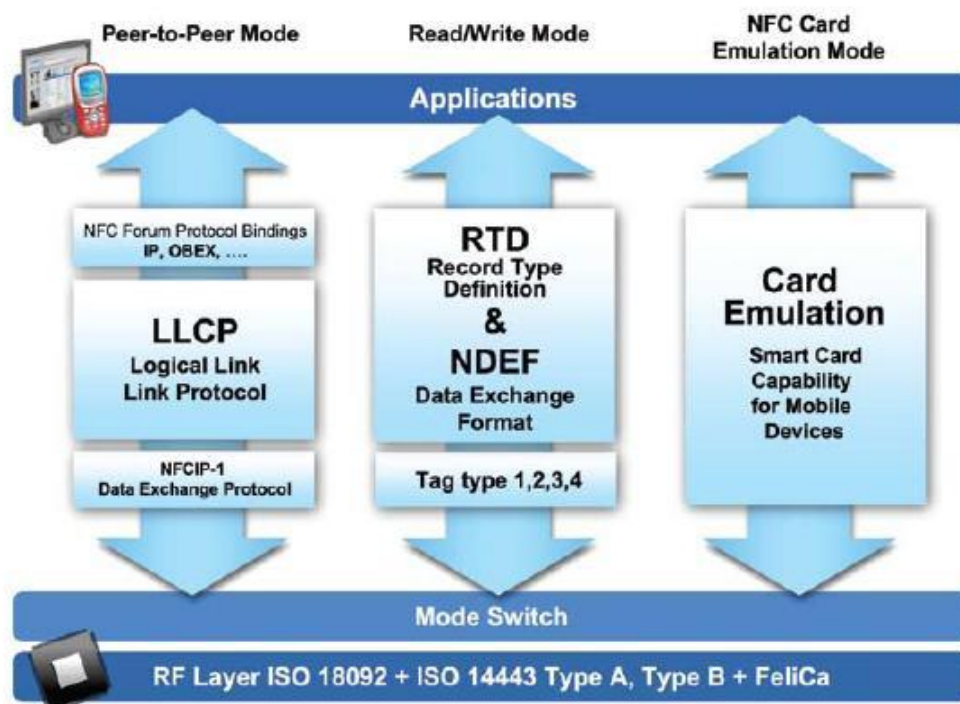


Imagem 2-2 - Modos de Operação NFC (8)

2.1.2. RFID

O RFID ou Radio-Frequency Identification (Imagem 2-3) é uma tecnologia que permite a troca de dados entre um leitor e uma etiqueta electrónica dum objecto a partir de ondas rádio.

A tecnologia RFID já existe desde a segunda guerra mundial mas só recentemente é que a tecnologia expandiu e começou a ser mais utilizada. O atraso na expansão deve-se ao facto da tecnologia ser muito cara para um consumo alargada e só agora começa a ter preços de fabrico aceitáveis. A RFID actualmente é aplicada para vários âmbitos tais como no pagamento a partir de telemóveis, nos transportes públicos (autocarro, metro, etc...), nos animais para identificação, em empresas para a monitorização de stocks e em bagagens de aeroportos.

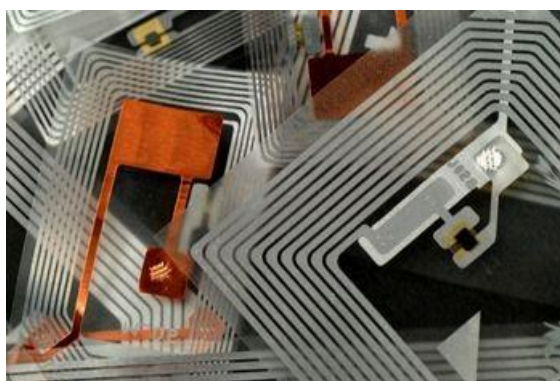


Imagem 2-3 - Etiquetas RFID (9)

Os sistemas RFID têm uma frequência de operação que pode estar dividida entre três bandas básicas: Baixa (125KHz), Alta (13,56 MHz) e Muito Alta (UHF). Um sistema básico é constituído por dois elementos: (10)

- **Etiquetas**, que consistem num microchip, que guarda os dados, e uma antena (elemento de acoplamento). Este elemento é fabricado de modo a que possa ser incorporado num objecto e tem um único número de série.
- **Leitores**, que podem ter uma ou mais antenas, emitem ondas rádio e recebem um sinal de resposta da etiqueta.

As etiquetas RFID podem ser de três tipos: Activas, Passivas ou Semi-Passivas. As etiquetas activas, que são substancialmente maiores que as passivas, contêm uma fonte de energia que as sustenta. Estas etiquetas podem ser acedidas até uma distância de 100 metros e conseguem receber sinais de baixa intensidade. Por outro lado as etiquetas passivas contêm

um chip ou circuito integrado para absorver as frequências rádio de um leitor e para enviar e receber os dados. Como não têm bateria, a energia é fornecida pelo leitor através das radiofrequências, ficando assim a etiqueta restrita ao campo do leitor. Para além disso, como estas etiquetas só conseguem transmitir sinais de baixa intensidade, precisam de estar muito próximas de etiquetas activas (menos de 5 metros) para receberem o sinal do leitor e puderem responder (11). As etiquetas semi-passivas funcionam da maneira similar às etiquetas passivas, sendo a diferença o facto de conterem uma bateria que possibilita a leitura de dados para distâncias maiores e permite á etiqueta funcionar de forma independente do leitor.

A tabela 2-1 apresenta um quadro resumo com os tipos de etiquetas e as suas características:

	RFID Activo	RFID Passivo
Energia	Bateria	Sem bateria interna
Força do sinal requerida	Baixa	Alta
Alcance de comunicação	Grande (+100m)	Pequeno (3m)
Armazenamento de dados	Grande capacidade	Pequena capacidade
Custo por Etiqueta	Geralmente, 15€ a 100€	Geralmente, 0,15€ a 5€
Tamanho da Etiqueta	Depende da aplicação	“Sticker” para cartão de crédito

Tabela 2-1- Diferenças entre etiquetas activas e passivas (12)

2.1.3. Comparação entre o NFC e o RFID

Tanto a tecnologia NFC como a RFID são muito semelhantes já que o NFC é uma extensão do RFID. A principal diferença é que o RFID é capaz de aceitar e transmitir para maiores alcances enquanto o NFC está restringido a poucos centímetros. No entanto o NFC permite a comunicação peer-to-peer e além disso permite a troca de dados entre dois dispositivos electrónicos. (Tabela 2-2)

	NFC	RFID
Tempo de setup	<0.1 ms	<0.1ms
Alcance	Até 10 cm	Até 3 m
Usabilidade	<ul style="list-style-type: none"> • Centrada no utilizador • Fácil • Intuitiva • Rápida 	<ul style="list-style-type: none"> • Centrada no objecto • Fácil
Selectividade	<ul style="list-style-type: none"> • Alta • Fixa • Segura 	<ul style="list-style-type: none"> • Parcialmente fixa

Casos de uso	<ul style="list-style-type: none">• Pagamentos• Obter acesso• Partilhar• Iniciar Serviços	<ul style="list-style-type: none">• Monitorização de objectos
--------------	--	---

Tabela 2-2 - Tabela comparativa NFC vs RFID (13)

2.2. Data mining

Data mining é o processo de extrair padrões de grandes quantidades de dados a partir da combinação de métodos estatísticos e de inteligência artificial com gestão de dados. Data mining tem vindo a evoluir como uma ferramenta importante no mercado moderno para a transformação de dados em informação inteligente dando uma vantagem informativa. Este processo é utilizado actualmente em várias áreas tais como:

- **Análises de mercado**
Para identificar os padrões de consumo do consumidor alfa. Os consumidores alfas são pessoas que desempenham uma papel essencial na relação com o conceito por detrás do produto, de seguida adoptando o produto, e finalmente validando para o resto da sociedade. A partir da análise dos dados recolhidos neste tipo de consumidores foi possível às companhias prever as tendências futuras de compra e projectas futuras procuras.
- **Ciência**
No estudo do corpo humano, um objectivo importante é de perceber as relações mapeadas entre as variações inter-individuais nas sequências do ADN humano e a variação da susceptibilidade a doenças. É uma ajuda importante para permitir melhorar os diagnósticos, para a prevenção e tratamento de doenças.
- **Educação**
Para estudar os factores que levam os estudantes a escolher comportamentos que reduzem a sua aprendizagem e para entender os factores que influenciam a retenção de estudantes na universidade.

2.2.1. Clustering

O clustering é um método de aprendizagem não-supervisionada que consiste na estruturação de um conjunto de observações em subconjuntos (clusters) de acordo com o seu grau de semelhança. O clustering é usado em vários campos, tais como na aprendizagem de máquinas, data mining, reconhecimento de padrões, análise de imagens e bioinformática.

O clustering pode ser de vários tipos, sendo eles:

- **Hierárquico**
O algoritmo pesquisa por clusters sucessivos usando para isso cluster estabelecidos anteriormente. Este tipo de algoritmo é normalmente aglomerativo (“bottom-up”) ou divisivo (“top-down”). Algoritmos

aglomerativos começam com cada ponto num cluster separado e junta-os em clusters sucessivamente maiores. Algoritmos divisivos começam com um conjunto e começam a dividi-lo em clusters mais pequenos.

- **Particional**

Estes algoritmos tipicamente determinam todos os clusters de uma só vez, mas também podem ser usados com algoritmos divisivos.

- **Densidade**

Estes algoritmos são designados para desenvolver clusters de tamanhos arbitrários. Nestes algoritmos o cluster é delineado quando a densidade na região onde esses se clusters se encontram ultrapassam um determinado limite.

- **Sub-espaço**

Este método pesquisa por cluster que apenas podem ser encontrados num determinado sup-espaço de dados. Este método ignora os atributos irrelevantes.

De seguida apresento alguns algoritmos que apresentem os tipos de clustering acima mencionados:

2.2.2. K-means clustering

O algoritmo k-means é um algoritmo particional que atribui cada ponto ao cluster com o centro mais perto. Os passos constituintes deste algoritmo são: (14)

1. Escolher o número de clusters, k.
2. Gerar aleatoriamente k clusters e determinar os centros dos clusters, ou directamente gerar k pontos aleatórios como centros do cluster.
3. Atribuir cada ponto ao centro do cluster mais próximo.
4. Recalcular o centro do novo cluster.
5. Repetir os dois passos anteriores até que um critério de convergência seja alcançado.

2.2.3. Cobweb

Cobweb é um algoritmo de cluster hierárquico que incrementalmente organiza as observações numa árvore de decisão. Cada nó na árvore contém um conjunto de probabilidades condicionais. Este método vai medir a diferença entre a probabilidade condicional de observar um produto dado e a probabilidade *a priori* de observar o valor de um atributo.

Cobweb considera quatro operações primárias (add, new, merge and split) que representam as maneiras possíveis de incorporar uma nova instância no nível de topo da hierarquia existente. Cobweb aplica cada operador e selecciona o que maximiza a utilidade da categoria da hierarquia resultante. (15)

2.2.4. DBScan

O DBScan (Density-Based Spatial Clustering of Applications with Noise) é um algoritmo de clustering do tipo densidade em que o número de clusters é descoberto automaticamente pelo algoritmo. Neste algoritmo os clusters são identificados analisando a densidade dos pontos. Regiões com alta densidade de pontos retratam a existência de clusters enquanto que as regiões com baixa densidade de pontos indicam clusters de ruído. Este algoritmo é particularmente adequado para grandes bases de dados, com ruído, e é capaz de identificar clusters com diferentes tamanhos (16).

2.2.5. EM Clustering

O EM Clustering (Expectation-Maximization Clustering) é um algoritmo de clustering do tipo particional que é frequentemente utilizado no clustering de dados para aprendizagem automática e em inteligência artificial. O algoritmo é composto por dois passos essenciais:

- Antecipação – Estima estatisticamente os valores das variáveis não observadas
- Maximização – Usando os valores estimados como verdadeiros, estima os parâmetros do modelo

3. Especificação e Definição do Sistema

3.1. Levantamento de Requisitos

Nesta secção irá ser feita uma descrição das funcionalidades que a aplicação deve realizar. No princípio os requisitos não estavam bem definidos pela empresa por isso houve um significativo trabalho inicial de discussão e análise de requisitos. Com a especificação de requisitos o leitor irá perceber melhor quais as funções e o comportamento que a aplicação deve desempenhar.

3.1.1. Casos de uso

Os casos de uso são uma descrição das interacções que ocorrem entre o utilizador e o sistema. Na Imagem 3-1 apresenta-se os casos de uso definidos que correspondem às funcionalidades que o sistema deve suportar. Os casos de uso também vão servir para, no capítulo de resultados/testes, verificar se os objectivos traçados foram cumpridos.

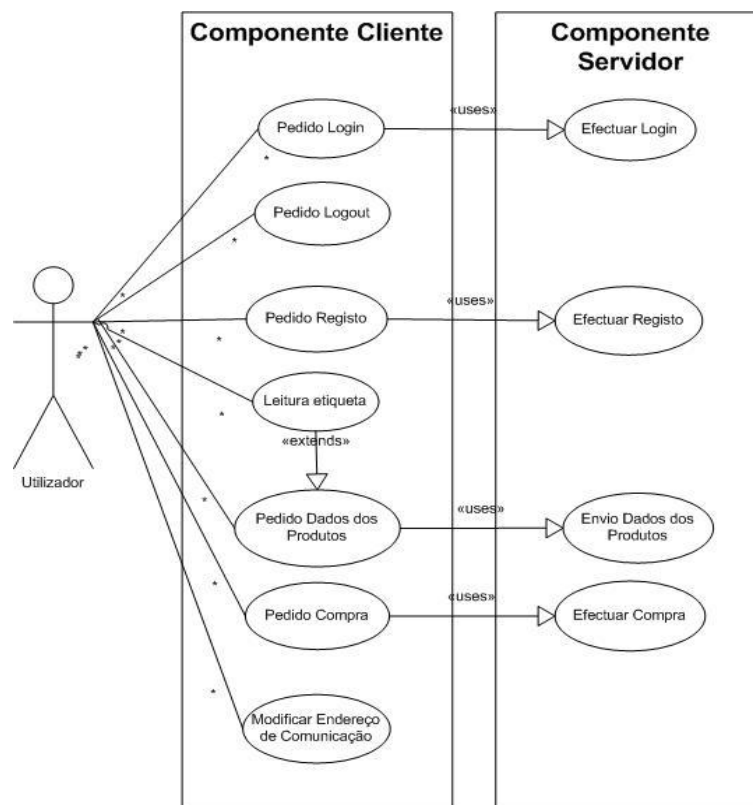


Imagem 3-1 - Casos de uso

Os casos de uso presentes na Imagem 4-1 são o de login, em que numa primeira fase o utilizador, na componente cliente, digita o seu username e password e de seguida essa informação é enviada para o componente servidor para verificar se está em concordância com o registado na base de dados e caso seja verdade o login é efectuado. O caso logout, que só diz respeito à componente cliente, em que o utilizador requisita o logout e que caso o utilizador tenha a autenticação activa, o logout é efectuado. No caso registo o utilizador preenche os dados requeridos para efectuar o registo na componente cliente e de seguida esses dados são enviados para o componente servidor e caso todos os campos estejam completos e não exista nenhum utilizador com o mesmo username o registo é efectuado. Outro dos casos de uso é o de leitura de etiqueta NFC, que consiste na leitura dos dados dum produto com etiqueta NFC por parte da componente cliente. O caso de pedido dos dados dos produtos acontece por consequência do caso de leitura dum etiqueta, em que após a leitura é feito, pela componente cliente, um pedido à componente servidor dos dados relativos ao produto lido assim como as sugestões de outros produtos. No caso de pedido de compra, o utilizador faz o pedido de compra dum produto por ele desejado, e é enviado esse pedido para o componente servidor que, caso o produto esteja presente na base de dados essa compra é validada e acrescentada ao histórico de compras do utilizador. No último caso de uso, modificar endereço de comunicação, é dada a possibilidade ao utilizador de modificar o endereço e porta com a qual comunica de maneira que possa estar em concordância com o endereço e porta do servidor.

3.2. Arquitectura do sistema

O sistema que permite efectuar os objectivos desejados é o que está representado na Imagem 3-1. O sistema começa a funcionar quando o utilizador aproxima o seu telemóvel dum leitor/etiqueta NFC. Nesse instante os dois componentes entram em comunicação por NFC e o leitor/etiqueta NFC retorna um ID (identificação) ao telemóvel. De seguida o telemóvel vai enviar esse ID, através de GSM/GPRS/3G/Wi-Fi, para um servidor Java que contém todas as funções desempenhadas pela aplicação, que irá retornar as informações relativas ao produto/etiqueta observado, assim como sugestões que possam ser relevantes para o utilizador. Para tornar as sugestões o mais apelativas (inteligentes) possíveis é obrigatório que o utilizador faça um registo prévio na aplicação de modo a ter um perfil pessoal.

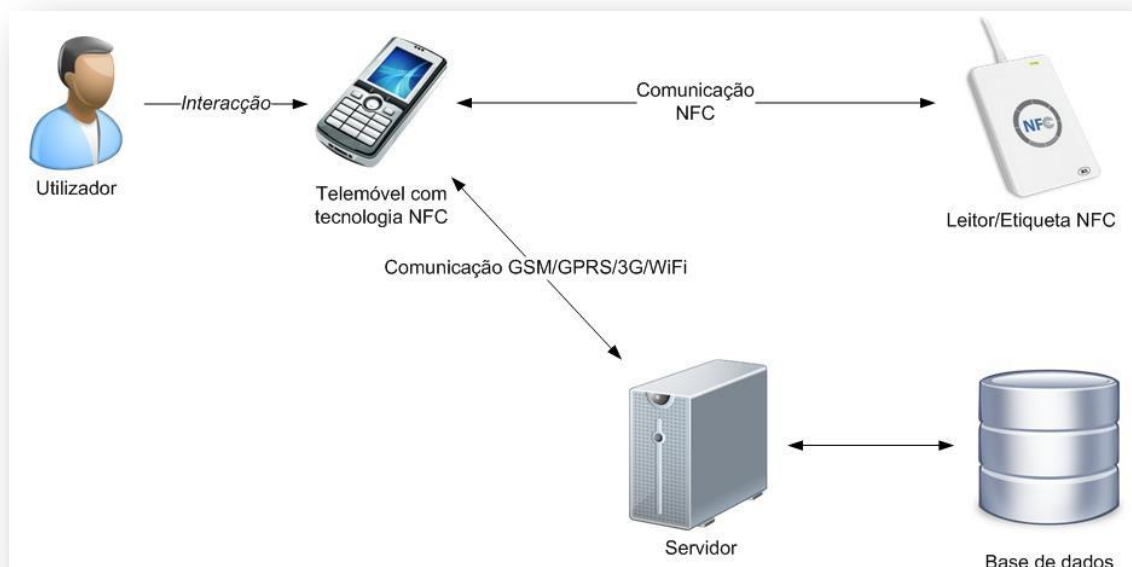


Imagem 3-2 - Arquitectura do sistema

Esta arquitectura, Imagem 3-2, é usualmente chamada de arquitectura de 3 níveis, em que os 3 níveis são a interface, a lógica e os dados. O nível interface é o nível mais alto e apresenta toda a parte visual do sistema. O nível de lógica é o nível abaixo do anterior e é o que controla as funcionalidades, processa comandos e faz decisões lógicas. Este nível também processa os dados dos dois níveis que o rodeiam. No nível de dados é onde está guardada toda a informação numa base de dados. A informação aqui contida é enviada para o nível de lógica e eventualmente para o utilizador. Este tipo de arquitectura é vantajoso porque permite que

qualquer um dos níveis seja modificado ou substituído sem ser necessário fazer modificações em nenhum dos outros níveis.

Pode-se dividir o sistema em duas componentes principais: o componente Cliente e o componente Servidor. A componente Cliente é composta pela aplicação no telemóvel enquanto a componente Servidor é composta pela aplicação Servidor e pela Base de dados.

3.2.1. Componente Cliente

A componente Cliente está presente na aplicação instalada no telemóvel e é onde estão implementadas as chamadas das funções da componente Servidor e a interface da aplicação. As interfaces da aplicação são:

- **Interface de Login** – Em que os utilizadores que já possuem registo fazem o seu login.
- **Interface de Registo** – Para os utilizadores ainda não registados fazerem o seu registo.
- **Interface do NFC** - Em que a aplicação está à espera que seja aproximada uma etiqueta NFC.
- **Interface do Produto e Sugestões** – Em que as características do produto aproximado assim como as sugestões correspondentes.

3.2.2. Componente Servidor

A componente Servidor é toda a parte da aplicação que não está integrada no telemóvel. Esta componente é constituída pelo Servidor, desenvolvido em Java, e pela Base de dados. Na Base de dados irão estar disponíveis todos os dados relativos aos utilizadores, aos produtos, e aos seus históricos de compra. Em relação ao Servidor as funcionalidades nele implementadas são:

- **Gestão dos utilizadores** - criação dos utilizadores e verificação da validade dos logins efectuados.
- **Gestão dos históricos** – Adição de novas compras feitas por utilizadores na Base de dados.
- **Características dos produtos** – Retorno das principais características do produto aproximado, acedendo à Base de dados, e das respectivas sugestões
- **Função algoritmo** – Em que a partir dos dados do utilizador e do produto por ele pesquisado, retorna sugestões o mais relevantes possíveis para o utilizador.

4. Desenvolvimento

4.1. Hardware utilizado

Para a definição do telemóvel para escolher era factor essencial que o equipamento tivesse suporte para NFC. O NFC é uma tecnologia que só recentemente começou a ser implementada em telemóveis de última geração e portanto não havia grande variedade de escolhas.

Após um período de pesquisa, chegou-se a três alternativas, o Nokia 6131, o Nokia C7-00 e o Google Nexus S (Imagem 4-1). O Nokia 6131 foi posto logo de lado porque, para além de ser um modelo de 2007, a sua interface é muito limitada e não possui a ligação a redes sem fios, funcionalidade importante para uma boa comunicação com o servidor. Após ponderação a escolha recaiu pelo Google Nexus S visto que utiliza o sistema operativo Android, sistema que está em expansão no mercado actual, e que serve-se da linguagem de programação Java. O Google Nexus S é o primeiro dispositivo com a mais recente plataforma Android, Gingerbread, a versão mais rápida do Android. O Gingerbread apresenta várias melhorias comparativamente com os anteriores, sendo as principais funcionalidades o suporte do NFC, um novo teclado e ferramenta de selecção de texto, internet (VOIP/SIP) calling, e com sensor giroscópio. Para além disso o Nexus S está dotado de um ecrã táctil Countour Display de 4 polegadas, de um processador Hummingbird de 1GHz e tem 16Gb de memória interna (17) .

Para além de um telemóvel com suporte NFC também são necessárias etiquetas NFC para testes, que já vêm incluídas na compra como acessórios e que servem perfeitamente para o ambiente académico em que se enquadra o trabalho.



Imagem 4-1 - Google Nexus S

4.2. Ferramentas/linguagens de programação

4.2.1. Eclipse IDE + Android SDK

O Eclipse é um software de desenvolvimento multi-linguagem que inclui um ambiente de desenvolvimento integrado (IDE). O software foi desenvolvido maioritariamente em Java e pode ser usado para a criação de programas em Java, C, C++, PHP, Python, entre outras. O Eclipse dá uso a complementos de modo a providenciar todas as funcionalidades, em contraste com outros softwares que já vem com as funcionalidades embutidas. O Eclipse vai ser usado em conjunto com o Android SDK, para permitir o desenvolvimento e uma melhor integração com o sistema operativo Android.

4.2.2. Xampp

O Xampp é um software livre para a criação de um servidor local, que contém uma base de dados MySQL, um servidor WebApache, e interpretadores para linguagens de programação PHP e Perl. O Xampp foi inicialmente desenvolvido para permitir aos programadores testar os seus trabalhos sem necessitarem de acesso à internet. Actualmente é usado como servidor de páginas Web e na criação e manipulação de base de dados MySQL.

4.2.3. WEKA

O Weka é um software que contém um pacote de ferramentas de data mining. O software tem como objectivo agregar algoritmos provenientes de diferentes paradigmas da inteligência artificial dedicada ao estudo de aprendizagem por parte das máquinas. O Weka encontra-se licenciado ao abrigo da GNU sendo portanto possível estudar e alterar o respectivo código.

Este software vai ser utilizado dando uso ao seu código-fonte, onde a partir dele vão ser chamadas funções para realizar o clustering dos dados.

4.2.4. Java

Linguagem de programação, que deriva grande parte da sua sintaxe no C/C++, orientada ao objecto. Esta linguagem é caracterizada pela sua compilação para bytecode, e não para código nativo como as linguagens convencionais. Isso faz com que uma das características principais do Java seja a portabilidade, já que qualquer programa desenvolvido

em Java funciona da mesma maneira independentemente do sistema operativo ou do hardware do dispositivo. O Java é actualmente uma das linguagens mais populares, e é amplamente utilizada em aplicações para Web.

Esta linguagem vai ser usada para o desenvolvimento da aplicação telemóvel e também para o desenvolvimento do servidor. O Eclipse vai ser a ferramenta utilizada para programar todo o código desenvolvido.

4.2.5. SQL

Linguagem de base de dados concebida para gerir dados em base de dados relacionais. Muitas das características originais do SQL foram inspiradas na álgebra relacional. Esta linguagem diferencia-se das outras linguagens de consulta de base de dados no sentido em que uma consulta SQL especifica a forma do resultado e não o caminho para lá chegar.

O SQL vai ser utilizado para a criação de novos dados (utilizadores, produtos, etc..) e também para a selecção de dados que estão alojados na base de dados. Todos os dados relativos á base de dados vão estar alojadas no servidor com auxílio da ferramenta Xampp.

4.3. Base de dados

4.3.1. Modelo de dados

A base de dados foi implementada em MySQL 5.0.7 e foi criada de acordo com o modelo de dados apresentado na Imagem 4-2. Como já foi referido anteriormente, a base de dados faz parte do componente servidor e foi criada localmente usando para isso o programa XAMPP. Na secção Anexo, na subdivisão Tabelas Base de Dados, são descritas pormenorizadamente as tabelas das bases de dados assim como a descrição do que cada dado compreende.

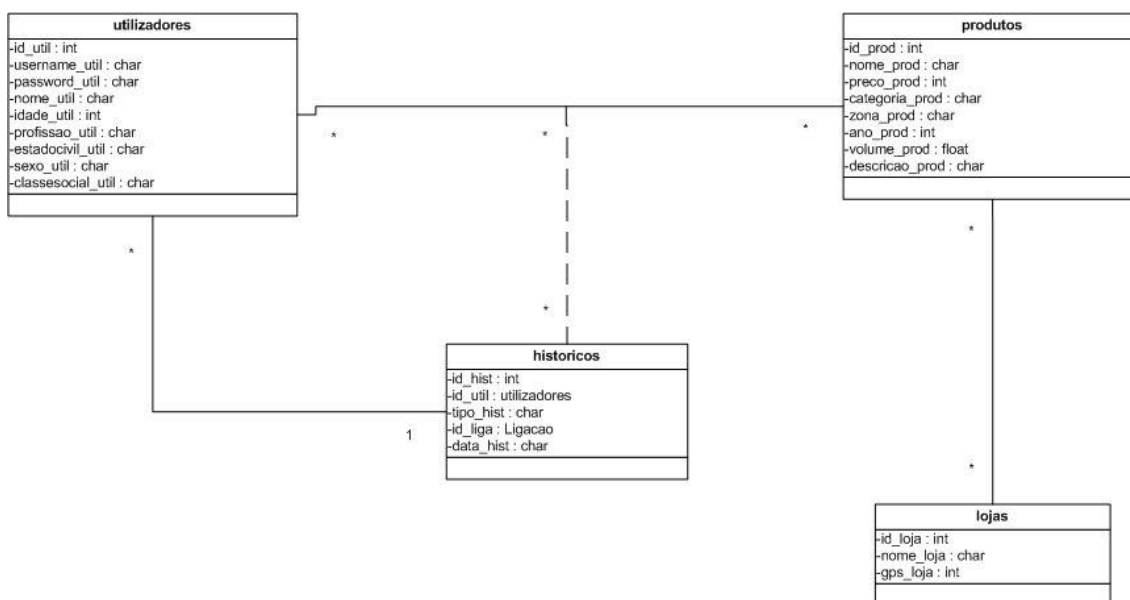


Imagem 4-2 - Modelo de dados

Como se pode perceber pela imagem acima, a base de dados é constituída por 5 tabelas, uma delas implícita na ligação entre as lojas e os produtos. Essas tabelas são as seguintes:

- **produtos** – Nesta tabela são apresentados os produtos presentes na base de dados. Em produtos guardam-se vários dados tais como, nome do produto, preço, URL da imagem do produto, entre outros.
- **utilizadores** – Tabela que guarda os dados relativos a cada utilizador. Esta tabela compreende vários dados, tais como o username, a password, o nome do utilizador, a idade, a profissão, entre outros.

- **historicos** – Tabela que contém as compras efectuadas pelos utilizadores. Para isso usa o ID do utilizador e o ID de ligação.
- **lojas** – Apresenta as lojas que existem na base de dados. Esta tabela foi criada para facilitar o trabalho em implementações futuras (explicado no Capítulo 5). Esta tabela contém o nome da loja e a posição GPS da loja.
- **ligacoes** – Compreende as disponibilidades presenciais dos produtos em determinadas lojas. Para isso usa o ID do produto e o ID da loja.

4.3.2. Dados iniciais

Inicialmente teve que ser criado uma base inicial em média escala porque para o clustering, e por arrastamento os testes efectuados no Capítulo 6, convém ter uma base de dados o maior possível para as sugestões obtidas pelo clustering serem o mais eficaz possível. Esta base teve que ser criada de modo a que utilizadores com perfis equivalentes tenham históricos de compras com pontos de contacto já que estes dados vão ser usados no clustering e assim ajuda a perceber melhor se as recomendações obtidas são as mais correctas. Como o clustering é efectuado aos perfis e ao histórico de compras foi dada mais ênfase na criação das tabelas relativas aos utilizadores e aos históricos.

Para os utilizadores foi criado um registo inicial de 100 utilizadores e criou-se de maneira a que os clusters criados sejam o mais equilibrados possível, ou seja que tenham praticamente o mesmo número de utilizadores. Para isso criou-se uma base de dados de utilizadores que apresenta 100 registos iniciais que contempla uma grande variedade de perfis, mas que apesar disso tenham vários dados em comum.

Em relação a tabela de produtos foi decidido, após conversação com a empresa CreativeSystems, que a tabela iria apenas contemplar apenas um tipo de produto e que ficou definido que seria vinho. Esta tabela irá contemplar dados relativos ao nome, volume, tipo de vinho (tinto, branco etc...), região de fabrico do vinho, preço, entre outros.

Relativamente aos históricos de compra foi criada uma tabela de preferências de compras induzidas (tabela 4-1) que contempla 1000 dados iniciais, 10 compras para cada utilizador, em que para cada caso possível de cada variável dos perfis dos utilizadores foram definidos produtos e depois para definir os históricos para cada utilizador teve-se em conta o maior número de produtos em comum para cada perfil. Por exemplo para um utilizador que tenha 20 anos, seja solteiro e estudante repete o produto com ID 1 e 4 duas vezes e o ID 2 e 3 uma vez, e na base de dados dos históricos para esse utilizador coloca-se mais compras do produto 1 e 4 do que dos produtos 2 e 3.

Idade	ID Produto	Estado Civil	ID Produto	Profissão	ID Produto
18-25	1,4	Solteiro	1,2,3	Advogado	12,13
26-40	2,3	Casado	6,7	Engenheiro	3,12
41-60	5	Divorciado	11	Estudante	4
61-∞	12,13	Viúvo	9	Médico	13
				Professor	7,8
				Motorista	9,10
				Secretário	1,2

Sexo	ID Produto	Classe Social	ID Produto
Masculino	4,7	Baixo	1,2,3
Feminino	2,9,11	Média	6,7
		Alta	9,10,13

Tabela 4-1 - Preferências do histórico

4.4. Interface do telemóvel

A interface do telemóvel, como já foi referido anteriormente, foi desenvolvida na componente cliente e tem como principal função permitir ao utilizador interagir com o sistema final. Esta interface foi desenvolvida usando o Eclipse + Android SDK, e foi desenvolvida em linguagem Java.

A interface do telemóvel vai ser constituída por 5 menus principais que são:

- Menu Login
- Menu Registo
- Menu Em Espera
- Menu Ver Produto
- Menu Opções

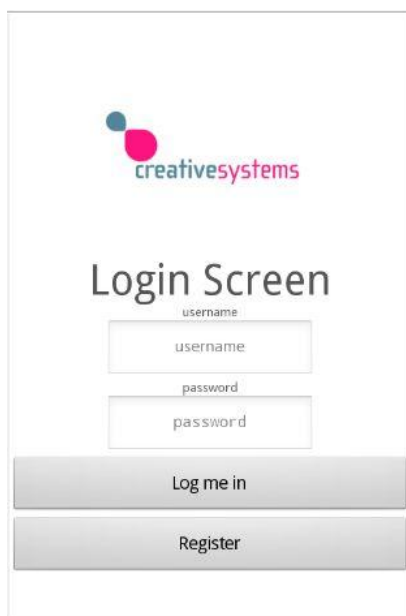


Imagem 4-3 - Menu Login



Imagem 4-4 - Menu Registo

No menu Login o utilizador tem a opção de fazer o login, se já tiver um registo no sistema, ou caso contrário é-lhe dado a opção de fazer um registo na aplicação.

No menu Registo o utilizador preenche o formulário com o username, password, nome, idade, profissão, estado civil, sexo e classe social. Após efectuar o registo o utilizador volta ao menu Login onde pode efectuar a autenticação com os dados do seu novo registo.

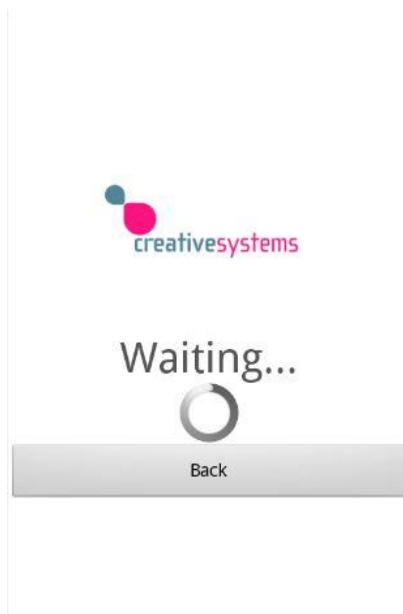


Imagem 4-5 - Menu Em Espera



Imagem 4-6 - Menu Ver Produto

No menu Em Espera a aplicação está à espera que um produto com etiqueta NFC seja aproximado e quando isso acontece salta para o menu Ver Produto.

No menu Ver Produto é apresentado o nome e preço do produto que o utilizador aproximou, assim como os dados das sugestões aconselhadas. O utilizador, se assim o pretender, pode comprar o produto pesquisado ou então seleccionar uma das sugestões e essa sugestão irá aparecer como produto principal e serão feitas novas sugestões em função do produto seleccionado. Neste menu, o utilizador pode aproximar outro produto que lhe será dadas as informações sobre ele e as sugestões, não sendo por isso necessário voltar ao menu Em Espera para pesquisar por outro produto.

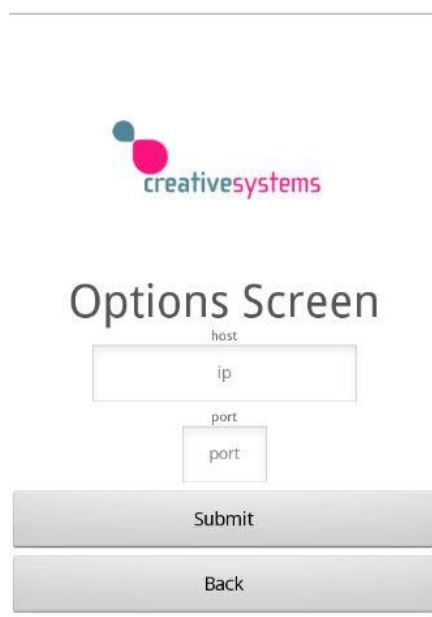


Imagem 4-7 - Menu Opções

O menu Opções é possível ser acedido a partir de todas as interfaces e permite ao utilizador modificar o endereço e a porta caso não esteja em concordância com o do servidor.

4.5. Protocolo de comunicação

Para a comunicação entre a componente cliente, telemóvel, e a componente servidor, base de dados e servidor, teve que ser criado um protocolo de comunicação. Durante a definição do protocolo tiveram que ser definidos comandos para serem utilizados durante as interações possíveis. Para cada interação foi criado um campo inicial para se saber qual a operação que se deseja realizar (Exemplo: login;username;password). As interações possíveis são as seguintes:

- **Operação Login**

- Descrição funcional: Através deste comando é verificado se o username e password associados ao utilizador estão correctos.
- Parâmetros de entrada: username e password.
- Parâmetros de saída: true/false.
- Exemplo:

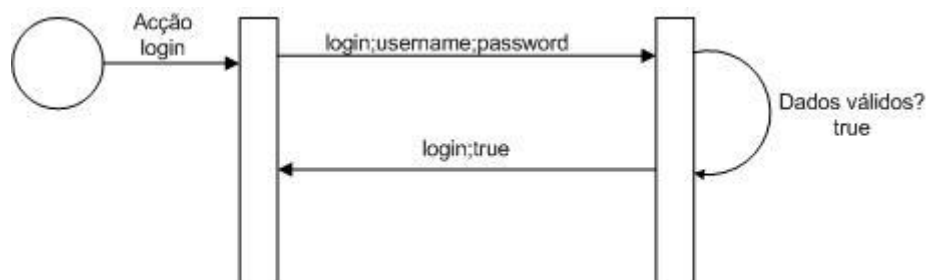


Imagem 4-8 - Exemplo operação Login

- **Operação Registo**

- Descrição funcional: Através deste comando o utilizador faz o novo registo de utilizador no sistema.
- Parâmetros de entrada: username, password, nome, idade, profissão, estado civil, sexo e classe social.
- Parâmetros de saída: true/false
- Exemplo:

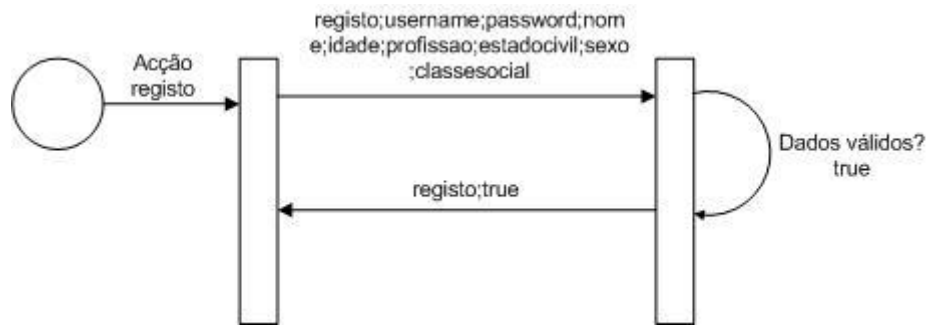


Imagem 4-9 - Exemplo operação Registo

- **Operação Buscar Produto**

- Descrição funcional: Através deste comando o servidor retorna a descrição do produto determinado pelo ID.
- Parâmetros de entrada: ID do produto.
- Parâmetros de saída: ID do produto, nome do produto, preço do produto e imagem do produto.
- Exemplo:

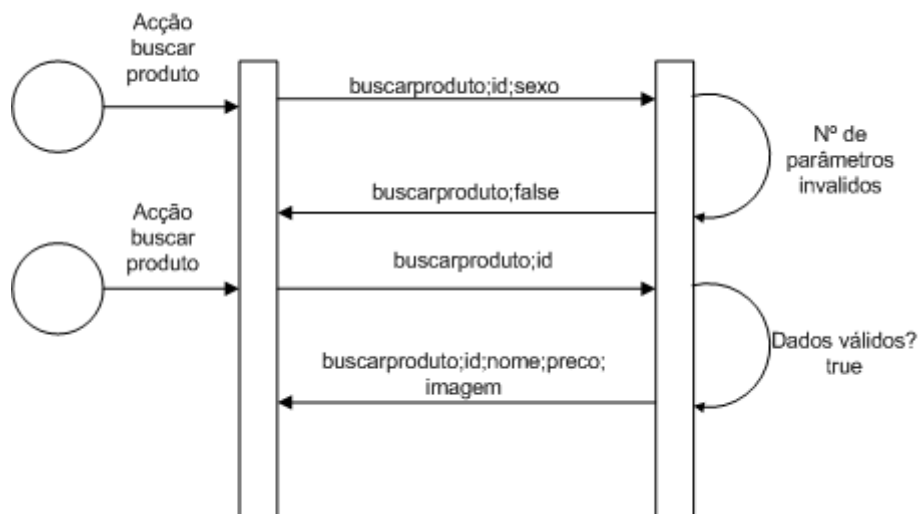


Imagem 4-10 - Exemplo operação Buscar Produto

- **Operação Sugestões**

- Descrição funcional: Através deste comando o servidor retorna as sugestões dos produtos dependendo do perfil do utilizador e do produto pesquisado.
- Parâmetros de entrada: username e ID do produto.
- Parâmetros de saída: ID do produto 1, nome do produto 1, preço do produto 1, imagem do produto 1, ID do produto 2, nome do produto 2, preço do produto 2, imagem do produto 2, ID do produto 3, nome do produto 3, preço do produto 3 e imagem do produto 3.

- Exemplo:

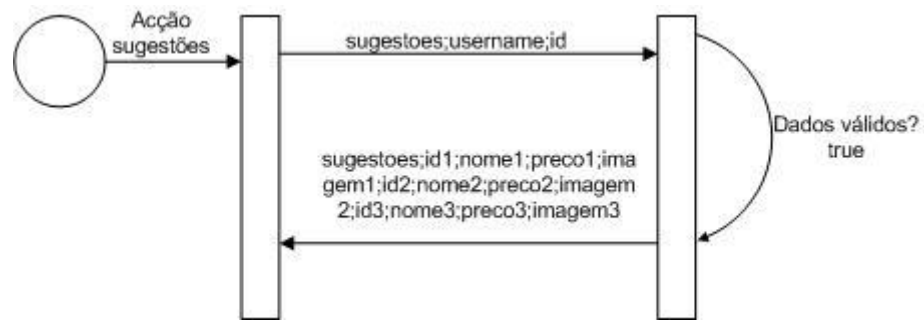


Imagem 4-11 - Exemplo operação Sugestões

- **Operação Compra**

- Descrição funcional: Através deste comando o utilizador efectua a compra do produto pesquisado.
- Parâmetros de entrada: username e ID do produto.
- Parâmetros de saída: true/false.
- Exemplo:

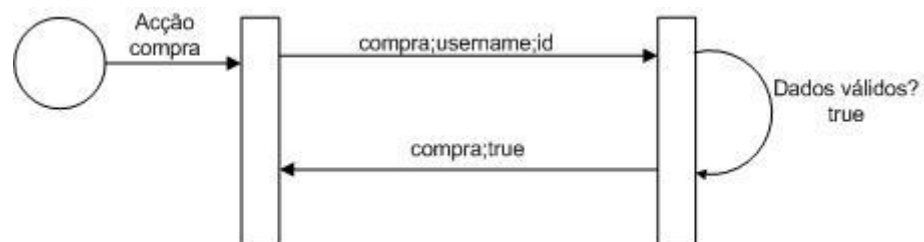


Imagem 4-12 - Exemplo operação Compra

Para além de prever o comportamento das operações quando estas decorrem correctamente também foram criadas respostas para quando nas entradas não entram valores correctos. Os erros a ser tratados são, por exemplo em qualquer operação se os campos de entrada não forem os previstos em termos de quantidade o servidor retorna false. Outra situação é no login em que é feita em comparação entre o username e a password digitados pelo utilizador estão de acordos com o que está descrito na base de dados e caso isso seja verdade retorna true, senão retorna false. Este passo é muito importante por assim cada utilizador fica com um perfil associado que mais tarde é usado para dar as sugestões. Como o foco principal desta dissertação passa pelo conceito de Ambiente Intelligente e por isso não foi dada grande ênfase à confidencialidade do utilizador. Apesar disso foi dada a possibilidade ao utilizador de mudar o endereço e porta de comunicação de maneira que fique de acordo com a do servidor de modo que só comunique com ele. Para além disso para cada pedido do utilizador enviada uma resposta que consiste num booleano, true ou false, nunca denunciando nem o username nem password do utilizador. Os dados de username e password inseridos no

telemóvel são apenas guardados localmente no telemóvel de modo que em futuras utilizações da aplicação não seja necessário estar a inserir novamente os dados de login. Se o utilizador assim o desejar pode efectuar logout sendo esses dados de login apagados embora isso traga como desvantagem o facto de na próxima utilização o utilizador ter de inserir o seu username e password.

4.6. Algoritmo

Para se obterem as sugestões inteligentes teve que ser utilizado um algoritmo que pudesse processar as decisões de modo a indicar as melhores escolhas para cada perfil de utilizador. Foi por isso decidido utilizar-se o software Weka e as suas bibliotecas de clustering. O clustering vai ser usado nos perfis dos utilizadores, onde foram definidas quais as variáveis (profissão, sexo, etc...) que deviam ser avaliadas e a partir daí os utilizadores ficam divididas em grupos (clusters), em que em cada grupo os utilizadores são o mais similares entre si e o mais diferente dos outros grupos. O algoritmo de clustering é efectuado em cada pesquisa do produto, ou seja se entre pesquisas for adicionado um novo utilizador o próximo clustering já irá ter em conta esse novo utilizador, funcionando portanto de forma autónoma.

Para além disso é feita uma busca dos utilizadores que mais compraram o produto a ser pesquisado pelo utilizador. Por exemplo, se o utilizador aproxima o produto X é feita uma pesquisa na base de dados dos utilizadores que mais compraram o produto X, e de seguida é realizada uma averiguação do seu histórico de compras e isso em conjunto com o clustering dá as sugestões ao utilizador.

5. Resultados/testes

5.1. Clustering

Para a definição do melhor clustering foram feitos alguns testes com o auxílio da aplicação Weka e das suas bibliotecas. Para realizar estes testes eram seleccionadas as variáveis que se pretendem que sejam avaliadas e de seguida são feitos vários tipos de clustering e são comparados os resultados obtidos. As variáveis que vão ser avaliadas fazem parte da tabela utilizadores e são a idade, a profissão, o estado civil, o sexo e a classe social.

Vão ser realizados dois tipos de teste para definir qual o melhor clustering a utilizar. O primeiro teste consiste na visualização do número de clusters que foram criados e na verificação se todas as instâncias de utilizadores estão distribuídas pelos clusters. No segundo teste vão ser avaliados as 3 principais sugestões de produtos que são obtidos para um determinado utilizador com um determinado perfil tendo em conta que cada variável (idade, profissão, sexo, etc...) tem uma preferência de produtos induzida como foi referido no Capítulo 4, secção de Dados Iniciais.

Numa primeira fase e através do primeiro teste, para se escolher qual o melhor clustering vai ser tomado em conta o número de cluster que são criados, já que cada cluster corresponde a uma aglomeração de tipos de perfis de utilizador com dados em comum e se houver muitos clusters irá significar que as sugestões dadas ao utilizador final serem de uma ou duas instâncias de utilizadores, tendo portanto uma menor probabilidade de ser relevantes para o utilizador. Também não é aconselhável obter muito poucos clusters porque isso iria significar que existiriam muitos utilizadores no mesmo cluster e as sugestões que iriam ser obtidas seriam de um grande aglomerado de utilizadores, e seriam portanto pouco relevantes. Se estes critérios forem preenchidos passa-se ao segundo teste em que vão ser comparadas as sugestões obtidas para um determinado utilizador com as respectivas preferências ditadas pelas variáveis presentes no seu perfil. Para sumarizar o que se pretende é portanto um clustering que apresenta um número razoável de clusters, em que as instâncias de utilizadores estejam dispersas pelos clusters e em que as sugestões obtidas estejam de acordo com o seu perfil.

O primeiro teste realizado foi com um clustering do tipo densidade, de nome DBScan, em que a definição do número de clusters é feita automaticamente. Os resultados obtidos foram os seguintes:

=== Model and evaluation on training set ===

DBScan clustering results

Clustered DataObjects: 100
Number of attributes: 4
Epsilon: 0.9; minPoints: 6
Index: weka.clusterers.forOPTICSAndDBScan.Databases.SequentialDatabase
Distance-type: weka.clusterers.forOPTICSAndDBScan.DataObjects.EuclidianDataObject
Number of generated clusters: 2
Elapsed time: ,03

Clustered Instances

0	8 (57%)
1	6 (43%)

Unclustered instances : 86

Imagem 5-1 – Clustering DBScan de 6 pontos

DBScan clustering results

Clustered DataObjects: 100
Number of attributes: 4
Epsilon: 0.9; minPoints: 2
Index: weka.clusterers.forOPTICSAndDBScan.Databases.SequentialDatabase
Distance-type: weka.clusterers.forOPTICSAndDBScan.DataObjects.EuclidianDataObject
Number of generated clusters: 25
Elapsed time: ,02
Unclustered instances : 18

Imagem 5-2 – Clustering DBScan de 2 pontos

Foram efectuados 2 testes com este tipo de clustering , no primeiro caso foi feito com um mínimo de 6 pontos em comum e no segundo com um mínimo de 2 pontos em comum. No primeiro caso apenas foram criados 2 clusters e obteve-se 86 instâncias de utilizadores que ficaram por ser seleccionados para um cluster. No segundo caso foram criados 25 clusters e houveram 18 instâncias de utilizadores que ficaram por serem seleccionados. Tanto num caso como noutro os resultados não foram satisfatórios já que se pretende que todos os utilizadores façam parte de um cluster. Por estas razões não se fez experiências com o segundo tipo de teste.

O segundo teste foi realizado com um clustering do tipo hierárquico, designadamente o Cobweb. Este tipo de cluster também define automaticamente o número de clusters. Os resultados obtidos foram:

```

=== Model and evaluation on training set ===

Number of merges: 33
Number of splits: 24
Number of clusters: 146

```

Imagem 5-3 - Clustering Cobweb

Apesar de neste clustering já não haver instâncias de utilizadores por seleccionar foram criados 146 clusters o que é demasiado pelos motivos já referidos acima na definição do que será um melhor tipo de clustering e por isso neste caso não se realizou o segundo tipo de teste.

De seguida foram realizados testes com 3 tipos de clustering em que se define *a priori* o número de clusters que se pretende obter. Para os 3 casos definiu-se o número de clusters igual a 5. Obteve-se os seguintes resultados:

```

=== Model and evaluation on training set ===

Clustered Instances

0      93 ( 93%)
1       2 (  2%)
2       3 (  3%)
3       1 (  1%)
4       1 (  1%)

```

Imagem 5-4 - Clustering Hierarchical de 5 clusters

```

=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 136.53419168991118
Missing values globally replaced with mean/mode

Clustered Instances

0      14 ( 14%)
1      23 ( 23%)
2      17 ( 17%)
3      26 ( 26%)
4      20 ( 20%)

```

Imagem 5-5 - Clustering k-means de 5 clusters

```
=== Model and evaluation on training set ===
```

```
EM
```

```
==
```

```
Number of clusters: 5
```

```
Clustered Instances
```

```
0      21 ( 21%)
1       6 (  6%)
2      34 ( 34%)
3       8 (  8%)
4      31 ( 31%)
```

Imagem 5-6 - Clustering EM de 5 clusters

No primeiro caso, Imagem 5-4, usou-se um clustering do tipo hierárquico de nome Hierarchical Clustering, enquanto no segundo e terceiro caso, Imagem 5-5 e Imagem 5-6, usou-se clustering do tipo particional, de nome K-means e Expectation-Maximization respectivamente.

Na comparação dos resultados exclui-se logo o Hierarchical Clustering porque embora tenha 5 clusters as instâncias de utilizadores estão concentradas no mesmo cluster (93%) ficando os outros clusters com poucas instâncias e por isso não se realizou o segundo tipo de teste para este clustering. Em relação ao K-means e ao Expectation Maximization ambos apresentam bons resultados em termos de distribuição de instâncias de utilizadores pelo clustering por isso vão ser realizadas experiências com o segundo tipo de teste.

Para este teste foi escolhido um utilizador aleatório da base de dados e vão ser verificadas as sugestões sugeridas com os dois clusterings. De seguida apresento a tabela 5-1 com o perfil do utilizador e as suas respectivas preferências induzidas, retiradas da tabela 4-1 do capítulo 4.

	Idade	Profissão	Estado Civil	Sexo	Classe social
	21	Professor	Solteiro	Masculino	Média
Preferências	1,4	7,8	1,2,3	4,7	6,7

Tabela 5-1- Preferências do utilizador 1

Nas tabelas seguintes apresento os resultados obtidos com o segundo tipo de teste para os clustering Expectation-Maximization e K-means com 5, 6 e 7 clusters:

	EM com 5 clusters	K-means com 5 clusters
Sugestões obtidas	ID: 4 Nome:Monsaraz Museu Aberto ID: 7 Nome:Bom Juiz Reserva ID: 9 Nome:Alianca Particular Bruto	ID: 7 Nome:Bom Juiz Reserva ID: 12 Nome:Artur Barros e Sousa Sercial ID: 3 Nome:Vinha DErvideira Branco

Tabela 5-2 - Sugestões obtidas para utilizador 1 com 5 clusters

	EM com 6 clusters	K-means com 6 clusters
Sugestões obtidas	ID: 4 Nome:Monsaraz Museu Aberto ID: 3 Nome:Vinha DErvideira Branco ID: 2 Nome:Castelar Bairrada Bruto ou Meio Seco	ID: 7 Nome:Bom Juiz Reserva ID: 6 Nome:Alvarinho Deu La Deu ID: 4 Nome:Monsaraz Museu Aberto

Tabela 5-3 - Sugestões obtidas para utilizador 1 com 6 clusters

	EM com 7 clusters	K-means com 7 clusters
Sugestões obtidas	ID: 4 Nome:Monsaraz Museu Aberto ID: 3 Nome:Vinha DErvideira Branco ID: 2 Nome:Castelar Bairrada Bruto ou Meio Seco	ID: 7 Nome:Bom Juiz Reserva ID: 4 Nome:Monsaraz Museu Aberto ID: 6 Nome:Alvarinho Deu La Deu

Tabela 5-4 - Sugestões obtidas para utilizador 1 com 7 clusters

Começando pela avaliação da tabela 5-2, percebe-se que os dois clusterings apresentam duas sugestões que estão de acordo com as preferências, para o EM o ID 4 e 7 e para o K-means o ID 7 e 3, em uma que não está, no EM o ID 9 e no K-means o ID 12. Nas tabelas 5-3 e 5-4 tanto para o clustering EM como para o clustering K-means as sugestões obtidas mantêm-se constantes, embora no K-means com 7 clusters haja uma mudança na ordem das sugestões. Outro resultado a que se chegue é que os dois clusterings agora apresentam sugestões que estão presentes nas preferências induzidas embora os resultados obtidos com o K-means sejam ligeiramente melhores já que as sugestões obtidas são sugestões que aparecem repetidas em variáveis diferentes mais vezes, o ID 7 aparece em 3 variáveis diferentes e o ID 4 aparece em 2 variáveis diferentes.

Para tirar dúvidas e se decidir qual o melhor clustering a aplicar foi realizado um segundo teste com outro utilizador, tabela 5-5.

	Idade	Profissão	Estado Civil	Sexo	Classe social
	45	Estudante	Casado	Feminino	Baixa
Preferências	5	4	6,7	2,9,11	1,2,3

Tabela 5-5 - Preferências do utilizador 2

Nas tabelas seguintes apresento os resultados obtidos para este utilizador usando os clustering Expectation-Maximization e K-means com 5, 6 e 7 clusters:

	EM com 5 clusters	K-means com 5 clusters
Sugestões obtidas	ID: 7 Nome:Bom Juiz Reserva ID: 13 Nome:Chateau Mouton Premier Grand Cru ID: 12 Nome:Artur Barros e Sousa Sercial	ID: 7 Nome:Bom Juiz Reserva ID: 4 Nome:Monsaraz Museu Aberto ID: 2 Nome:Castelar Bairrada Bruto ou Meio Seco

Tabela 5-6 - Sugestões obtidas para utilizador 2 com 5 clusters

	EM com 6 clusters	K-means com 6 clusters
Sugestões obtidas	ID: 7 Nome:Bom Juiz Reserva ID: 4 Nome:Monsaraz Museu Aberto ID: 13 Nome:Chateau Mouton Premier Grand Cru	ID: 4 Nome:Monsaraz Museu Aberto ID: 7 Nome:Bom Juiz Reserva ID: 2 Nome:Castelar Bairrada Bruto ou Meio Seco

Tabela 5-7 - Sugestões obtidas para utilizador 1 com 6 clusters

	EM com 7 clusters	K-means com 7 clusters
Sugestões obtidas	ID: 7 Nome:Bom Juiz Reserva ID: 13 Nome:Chateau Mouton Premier Grand Cru ID: 4 Nome:Monsaraz Museu Aberto	ID: 4 Nome:Monsaraz Museu Aberto ID: 7 Nome:Bom Juiz Reserva ID: 2 Nome:Castelar Bairrada Bruto ou Meio Seco

Tabela 5-8 - Sugestões obtidas para utilizador 1 com 7 clusters

Para este utilizador, no caso do clustering EM com 5 clusters obteve-se uma sugestão, ID 7, em comum com as das variáveis do utilizador enquanto para 6 e 7 clusters obteve-se duas sugestões similares, ID 4 e ID 7, à das preferências. Em relação ao K-means obteve-se sempre os mesmos valores, independentemente do número de clusters, embora troquem de ordem. Todas as sugestões obtidas por este clustering estão em concordância com as preferências induzidas.

Pode-se portanto concluir que o clustering K-means é mais eficaz para esta base de dados, estando as sugestões dadas de acordo com as preferências induzidas.

Já que no caso do clustering K-means as sugestões obtidas foram sempre semelhantes independentemente do número de clusters foram feitos uns testes finais em que se varia o número de clusters definidos para definir com quantos clusters deve o algoritmo funcionar. Os resultados obtidos apresentam-se nas imagens seguintes:

```
=== Model and evaluation on training set ===
```

```
kMeans  
=====
```

```
Number of iterations: 4  
Within cluster sum of squared errors: 137.87928052346888  
Missing values globally replaced with mean/mode
```

```
Clustered Instances
```

0	13 (13%)
1	19 (19%)
2	16 (16%)
3	14 (14%)
4	15 (15%)
5	23 (23%)

Imagem 5-7 - Clustering K-means com 6 clusters

```
=== Model and evaluation on training set ===
```

```
kMeans  
=====
```

```
Number of iterations: 5  
Within cluster sum of squared errors: 129.90730651019206  
Missing values globally replaced with mean/mode
```

```
Clustered Instances
```

0	13 (13%)
1	14 (14%)
2	14 (14%)
3	13 (13%)
4	15 (15%)
5	19 (19%)
6	12 (12%)

Imagem 5-8 - Clustering K-means com 7 clusters

```
=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 4
Within cluster sum of squared errors: 124.98614081028234
Missing values globally replaced with mean/mode

Clustered Instances

0      12 ( 12%)
1      11 ( 11%)
2      12 ( 12%)
3      13 ( 13%)
4      13 ( 13%)
5      21 ( 21%)
6       9 (  9%)
7       9 (  9%)
```

Imagem 5-9 - Clustering K-means com 8 clusters

Como se pode verificar pelas imagens acima, até 7 clusters as instâncias dos utilizadores estão cada vez mais dispersas pelos clusters. No caso com 8 clusters já não é isso que acontece, ficando o cluster 5 com uma maior percentagem de utilizadores que no caso com 7 clusters. O que isto significa é que com 8 clusters, as sugestões obtidas não vão ser tão eficazes já que se começam a formar grupos maiores de utilizadores. Chegou-se por isso à conclusão que a melhor solução é o algoritmo K-means com 7 clusters.

5.2. Validação dos casos de uso

Para ser feita uma avaliação qualitativa do sistema final foram feitos testes de integração. Estes testes vão permitir validar o devido funcionamento dos casos de uso.

5.2.1. Gestão do Login

O objectivo deste teste é verificar se os dados (username e password) inseridos pelo utilizador são correctamente lido pela aplicação, se esses dados são correctamente enviados ao servidor e se o servidor verifica correctamente a base de dados para confirmar se esses dados estão de acordo ou não com o presente na base de dados, respondendo true e false respectivamente.

Este teste atingiu todas as metas, já que a resposta dada pelo servidor foi sempre a esperada dependendo dos dados inseridos pelo utilizador.

5.2.2. Efectuar Logout

É solicitado com este teste verificar se, após o login ser efectuado, o logout é correctamente efectuado independentemente da interface em que a aplicação se encontra.

Este teste foi efectuado correctamente, não havendo erros a relatar.

5.2.3. Gestão do Registo

Neste teste é pretendido verificar se o utilizador ao não inserir dados nos campos todos o registo não é efectuado, e se é possível efectuar um registo com o mesmo username de outro utilizador já presente na base de dados. Para além disso este teste serve para verificar se, quando todos os dados são preenchidos correctamente, são enviados os dados correctamente para o servidor e se é inserido um novo utilizador na base de dados.

Foram feitas diversas inserções na base de dados de novo utilizadores e não existem erros a relatar. Também foram criados registos com campos em branco e a aplicação reagiu como esperado não deixando que o registo fosse efectuado e pedindo ao utilizador para

preencher os campos que tinha deixado em branco. Para testar o username foram criados registos com usernames que já existam na base de dados e o registo também não é efectuado.

5.2.4. Leitura da Etiqueta NFC

É pretendido com este teste que seja feita a leitura duma etiqueta NFC, nomeadamente dos dados relativos ao ID do produto.

Este teste foi realizado várias vezes não havendo problemas na leitura das etiquetas.

5.2.5. Gestão dos dados dos produtos

Este teste consiste no envio dos dados necessários do produto seleccionado para o componente servidor e, após passar por funções algorítmicas, no retorno para a aplicação do telemóvel dos respectivos dados do produto e das sugestões aconselhadas.

Foram realizados vários testes e foi sempre retornado o dados do produto e respectivas sugestões. Só não retornava dados quando o produto seleccionado não estava presente na base de dados, como era o esperado.

5.2.6. Efectuar Compra

O objective deste teste é verificar se o utilizador ao efectuar uma compra, os dados do produto que ele pretende comprar e do utilizador são correctamente enviados ao componente servidor e se os dados são inseridos com sucesso na base de dados.

Este teste foi realizado com sucesso sendo os dados enviados correctamente para o servidor e sendo os mesmos correctamente inseridos na base de dados.

5.2.7. Gerir Endereço de Comunicação

Com este teste é pretendido verificar se é possível modificar o endereço e a porta na aplicação telemóvel de modo a ficar em conformidade com o servidor e ser possível ser efectuada a comunicação entre os dois.

Para se verificar este caso foi modificado o endereço e porta do servidor de modo a que não fosse possível a comunicação com a aplicação do telemóvel e de seguida foi modificado os valores do endereço e porta na aplicação sendo após essa mudança possível a troca de dados entre os dois. Este teste foi realizado várias vezes sendo possível, após a mudança de endereço e porta na aplicação do telemóvel, a comunicação com o servidor, como era o pretendido.

6. Conclusões

6.1. Conclusão

Este trabalho tinha como objectivo a criação de uma aplicação para um telemóvel com tecnologia NFC que permitisse ao utilizador aproximar o dispositivo dum produto com etiqueta NFC e retornasse informações sobre esse produto e assim como sugestões que possam ser do agrado do utilizador. Para isso o utilizador teria de estar registado na aplicação para se poder fazer um perfil do mesmo e assim dar as sugestões o mais inteligentes possíveis ao olhar do utilizador.

O primeiro passo no desenvolvimento foi a definição da arquitectura do sistema, em que se definiu que haveria duas componentes, componente cliente e componente servidor. Na componente cliente iria ficar a aplicação do telemóvel enquanto no componente servidor ficaria a aplicação que teria o algoritmo de clustering e a base de dados de utilizadores e produtos.

Teve que ser criada uma base de dados inicial de utilizadores para ser criar cluster de perfis de utilizadores de modo a dar as sugestões inteligentes. Para além disso teve que ser criada uma base de dados iniciais de produtos e um histórico de compras dos utilizadores também para ajudar na determinação das sugestões dadas ao utilizador. De seguida foi criada o componente cliente que consistia na interface do telemóvel, na comunicação por NFC com etiquetas e na comunicação com o componente servidor tendo em conta um protocolo de comunicação. O protocolo de comunicação permite á componente cliente interagir com o componente servidor e retornar dados do mesmo. Para finalizar desenvolveu-se a aplicação servidor que conteria todas as funções que a aplicação telemóvel necessitaria. Entre elas está a função que decidirá quais as sugestões a dar ao utilizador, que para além de ver o histórico de compras dos utilizadores que compraram o produto que foi pesquisado pelo utilizador, também faz um clustering dos utilizadores e junta os utilizadores em clusters com perfis semelhantes.

Para concluir, e tendo em conta o já foi dito, posso afirmar que todos os objectivos propostos inicialmente foram alcançados com sucesso.

6.2. Implementações Futuras

Apesar de todos os objectivos terem sido cumpridos há sempre melhorias que podem ser feitas de modo a tornar a aplicação mais completa e mais apelativa para o utilizador. De seguida apresentarei funcionalidades futuras que poderiam trazer um valor acrescentado a esta aplicação.

Uma das funcionalidades seria acrescentar um sistema de rating dos produtos, que permitisse aos utilizadores avaliarem os produtos. Poderia se depois usar esses valores dos ratings para também auxiliar a tomada de decisão de quais as sugestões a apresentar.

Algum desenvolvimento também poderia ser acrescentado ao histórico de modo a que fosse não só possível adicionar as compras efectuadas mas também as visualizações efectuadas aos produtos, visto que o utilizador visualizar um produto pode ser um outro factor importante para auxiliar nas sugestões a apresentar.

Também poderia ser adicionada a aplicação a tecnologia GPS. Ou seja, ter em conta a posição em que o utilizador se encontra para só lhe sugerir produtos que se encontram em lojas na redondeza. A base de dados actual já está pronta para trabalhar com esta funcionalidade, visto que na tabela lojas já tem um parâmetro que contém as posições GPS das lojas.

Outra das funcionalidades possíveis seria a ligação a uma carteira virtual. O utilizador inseriria os seus dados do cartão de crédito e assim poderia efectuar uma “compra verdadeira” de um produto. Para esta funcionalidade funcionar teria que ser desenvolvida e aperfeiçoada a segurança da aplicação.

Para finalizar uma funcionalidade natural que podia ser desenvolvida seria o aumento da base de dados de utilizadores e o número de atributos que são avaliados pelo clustering.

7. Bibliografía

1. **Cassidy, Ruth.** *Call For Entries Touching the Future: NFC Forum Global Competition.* 2007.
2. **Paus, Annika.** *Near Field Communication in Cell Phones.* 2007.
3. **Forum, NFC.** *The Keys to Truly Interoperable Communications.* 2007.
4. **José Bravo, Ramón Hervás, Gabriel Chavira, Salvador W. Nava and Vladimir Villarreal.** *From Implicit to Touching Interaction: RFID and NFC Approaches.* 2008.
5. **Gabriel Chavira, Salvador W. Nava, Ramón Hervás, José Bravo, Carlos, Sánchez.** *Towards Touching Interaction: A Simple Explicit Input.*
6. **Chavira, G., Nava, S. W., Hervás, R., Villarreal, V., Bravo, J., Martín, S., Castro, M.** *Services through NFC technology in Aml Environment.* ERPAS 2008.
7. **Forum, NFC.** NFC Forum: Frequently Asked Questions. *NFC Forum.* [Online] <http://www.nfc-forum.org/resources/faqs/>.
8. **NFC, Forum.** NFC Forum: Forum Marketing White Paper. *Forum NFC.* [Online] http://www.nfc-forum.org/resources/white_papers/nfc_forum_marketing_white_paper.pdf.
9. **Buzzle.com.** Passive vs Active RFID Tags. *Buzzle.com.* [Online] <http://www.buzzle.com/articles/passive-vs-active-rfid-tags.html>.
10. **Gabriel Chavira, Salvador W. Nava, Ramón Hervás, José Bravo, Carlos Sánchez.** *Combining RFID and NFC Technologies in an Aml Conference Scenario.*
11. **McCormick, Meredith.** Active Vs. Passive RFID Tags (Radio Frequency Identification). *associatedcontent.com from Yahoo.* [Online] http://www.associatedcontent.com/article/2971594/active_vs_passive_rfid_tags_radio_frequency.html.
12. **solutions, atlas RIFD.** Active vs Passive. *atlas RFID solutions.* [Online] <http://www.atlasrfid.com/Technology/ActivevsPassive.aspx>.
13. **Oracle.** Oracle. *An Introduction to Near-Field Communication and the Contactless Communication API.* [Online] <http://java.sun.com/developer/technicalArticles/javame/nfc/>.
14. **Wikipedia.org.** Cluster analysis. *Wikipedia.org.* [Online] http://en.wikipedia.org/wiki/Cluster_analysis.
15. **Kiri Wagstaff, Claire Cardie.** *Clustering with Instance-level Constraints.* 2000.
16. **Adriano Moreira, Maribel Y. Santos and Sofia Carneiro.** *Density-based clustering algorithms – DBSCAN and SNN.* 2005.
17. **Inc., Google.** Nexus S. [Online] <http://www.google.pt/nexus/#/features>.
18. **Jantunen, Iiro.** *Architecture Development for Mobile-phone-centric Ambient Intelligence Applications.* 2010.

ANEXO I – TABELAS BASE DE DADOS

Tabela	produtos	
Descrição	Tabela que contém as informações dos produtos presentes na base de dados.	
Coluna	Tipo	Descrição
id_prod	bigint(20)	ID do produto.
nome_prod	varchar(150)	Nome do produto.
preço_prod	varchar(150)	Preço do produto.
categoria_prod	varchar(150)	Categoria/tipo do produto.
zona_prod	varchar(150)	Zona onde o produto foi fabricado.
ano_prod	varchar(150)	Ano do produto.
volume_prod	varchar(150)	Volume do produto.
imagem_prod	varchar(150)	Link para a imagem do produto.

Tabela	utilizadores	
Descrição	Tabela que contém o username e password dos utilizadores e seus respectivos perfis.	
Coluna	Tipo	Descrição
id_uti	bigint(20)	ID do utilizador.
username_util	varchar(30)	Username do utilizador.
password_util	varchar(30)	Password do utilizador.
nome_util	varchar(30)	Nome do utilizador.
idade_util	int(11)	Idade do utilizador.
profissão_util	varchar(30)	Profissão do utilizador.
estadocivil_util	varchar(30)	Estado civil do utilizador.
sexo_util	varchar(30)	Sexo do utilizador.
classesocial_util	varchar(30)	Classe social do utilizador.

Tabela	historicos	
Descrição	Tabela que contém os históricos de compras de todos os utilizadores.	
Coluna	Tipo	Descrição
id_hist	bigint(20)	ID do histórico.
tipo_hist	varchar(30)	Informação sobre o tipo de histórico.
id_util	int(11)	ID do utilizador.
id_liga	int(11)	ID de ligação.

Tabela	lojas	
Descrição	Tabela que contém as lojas presentes na base de dados.	
Coluna	Tipo	Descrição
id_loja	bigint(20)	ID da loja.
nome_loja	varchar(30)	Nome da loja.
gps_loja	varchar(30)	Posição GPS da loja.

Tabela	ligacoes	
Descrição	Tabela que contém as ligações entre os produtos e as lojas, ou seja apresenta a disponibilidade ou não de determinado produto em determinada loja.	
Coluna	Tipo	Descrição
id_liga	bigint(20)	ID da ligação.
id_prod	int(11)	ID do produto.
id_loja	int(11)	ID da loja.