**NÍCOLAS HENRIQUE VIEIRA TOLEDO**

**SEBASTIÃO BATISTA DE ANDRADE NETO**

**ESTATÍSTICAS POLÍTICAS**

**EM TEMPO REAL COM BIG DATA**

**UNIVERSIDADE DO VALE DO SAPUCAI**

**POUSO ALEGRE**

**2015**

**NÍCOLAS HENRIQUE VIEIRA TOLEDO**

**SEBASTIÃO BATISTA DE ANDRADE NETO**

**ESTATÍSTICAS POLÍTICAS**

**EM TEMPO REAL COM BIG DATA**

Pré-projeto de pesquisa apresentado à disciplina de TCC 1 do Curso de Sistemas de Informação como requisito parcial para obtenção de créditos.

**UNIVERSIDADE DO VALE DO SAPUCAI**

**POUSO ALEGRE**

**2015**

# INTRODUÇÃO

Nos dias atuais, vê-se um grande volume de dados circulando por todo o globo diariamente. Para mais fácil entendimento, foi dado o nome de “Big Data” a esse grande volume de dados. Porém a questão é: “O que fazer com Big Data?”.

De fato, esse volume de dados ocupa uma grande demanda TI de uma organização. Porém os benefícios adquiridos podem dar uma severa vantagem competitiva em relação aos concorrentes.

Redes Sociais como o Facebook, por exemplo, podem ser um diferencial em determinadas áreas, devido à grande quantidade de dados gerados diariamente.

Analisando em um cenário nacional, mais precisamente no âmbito político, os presidenciáveis Dilma Rousseff e Aécio neves, juntos tiveram cerca de 1.585.369 de menções no Facebook e 2.715.438 de Tweets no ano de 2014.

Como a disputa à presidência estava acirrada, marqueteiros perceberam a importância do imenso volume de dados gerados pelas redes sociais e bombardearam os eleitores com informações quase quem em tempo real dos candidatos, além é claro, de acusações ao concorrente. Redes Sociais tem uma vantagem consideravelmente interessante: Elas encurtam a relação com o eleitor.

Muitos usuários demonstram o contentamento com seus representantes através de redes sociais, pelo simples fato delas serem uma forma fácil de expressão. Geralmente, a população tente a se expressar quando ocorrem eventos de grandes proporções nacionais, como uma eleição presidencial. Neste ponto nasce o Big Data.

Pelo simples fato dessa grande quantidade de dados conter as opiniões mais diretas da população, estamos propondo um novo modelo de análise de opinião pública no Brasil: Estatísticas políticas em tempo real, usando Big Data.

O contexto da proposta é relativamente simples: Descobrir como a população enxerga as atitudes do governo, em tempo real.

Com essa análise de opinião pública, os políticos poderão saber se estão cumprindo ou falhando em seu dever.

Portanto, esse sistema de análise estatística é especificamente desenvolvido para auxiliar os nossos representantes a descobrir as carências da população, melhorando assim a qualidade de vida do povo.

## Objetivo Geral

Desenvolver um site que mostre informações em tempo real de opiniões políticas dos usuários de redes sociais.

## Objetivos Específicos

Os nossos objetivos específicos tem como base:

* Coletar uma grande quantidade de dados políticos de redes sociais através de palavras-chave;
* Coletar os dados;
* Analisar os dados de forma eficiente e rápida;
* Apresentá-los em tempo real em um site.

## Justificativas

Dentre tantos temas, foi escolhida a política por ser um fator presente na vida de todos os brasileiros. A intenção é ajudar os políticos a ajudar a população, informando a eles onde e como a população é carente.

Em relação opinião pública, foram escolhidas as redes sociais por serem atualmente a forma mais ampla de protesto político no Brasil.

E, em relação, à tecnologia, foi escolhido o Big Data por ser um assunto de atual relevância e diferencial competitivo no mercado mundial.

# FUNDAMENTAÇÃO TEÓRICA

## Big Data

A princípio, podemos definir o conceito de Big Data como sendo conjuntos de dados extremamente amplos e que, por este motivo, necessitam de ferramentas especialmente preparadas para lidar com grandes volumes, de forma que toda e qualquer informação nestes meios possa ser encontrada, analisada e aproveitada em tempo hábil.

De maneira mais simplista, a ideia também pode ser compreendida como a análise de grandes quantidades de dados para a geração de resultados importantes que, em volumes menores, dificilmente seriam alcançados. (Alecrim; Emerson, 2013).

## Banco de dados NoSQL

NoSQL é um movimento que promove soluções de armazenamento de dados não relacionais. Ele é composto por diversas ferramentas que, de forma particular e específica, resolvem problemas como tratamento de grandes volumes de dados, execução de consultas com baixa latência e modelos flexíveis de armazenamento de dados, como documentos XML ou JSON.

NoSQL não têm como objetivo substituir os bancos de dados relacionais, mas apenas propor algumas soluções que em determinados cenários são mais adequadas. Desta forma é possível trabalhar com tecnologias NoSQL e banco de dados relacionais dentro de uma mesma aplicação. (Java Magazine, 2013).

Os tipos de bancos de dados NoSQL que são mais usados são:

* Chave-Valor: Armazenam objetos indexados por chaves, e possibilitam sua busca a partir de sua respectiva chave.
* Documentos: Conjunto de documentos no formato JSON. Os documentos são tratados como objetos únicos, os quais possuem campos com os respectivos valores.
* Coluna: Formado por colunas que contem um conjunto de informações, semelhante a uma tabela.
* Grafos: Formado por dados distribuídos em forma de vértices e arestas, os quais possuem atributos tanto nas arestas quanto nos vértices.

## Hadoop

A definição de Open Source e Software Livre são quase equivalentes, uma vez que Software livre é Open Source, mas nem todo Open Source é Software Livre.

Segundo Silveira (2004, p.15) “Software Livre é Open Source. Open Source é um software que possui o código-fonte aberto. Entretanto é possível que um software de fonte aberta não assegure as quatro liberdades que caracterizam o software livre.”

De acordo com Alves (2005), o movimento de código aberto teve inicio em 1997 por pessoas que trabalhavam com Software Livre, e que se sentiam incomodadas com o caráter ideológico da FSF e a licença do GNU GPL.

De acordo com Alves (2004 apud Perens 1999) o Open Source atende há alguns critérios:

* O programa deve poder ser redistribuído sem a exigência de pagamentos de qualquer espécie.
* Todos os programas devem incluir o código-fonte. Caso não seja distribuído com o código- fonte, deve haver indicações claras de como obtê-lo.
* A licença deve permitir modificações e trabalhos derivados que, devem poder ser distribuídos nos mesmos termos de licença original do software.
* A licença pode restringir de trabalhos cujo original seja diretamente modificado, desde que garanta a distribuição de arquivos de correção que possam ser aplicados no momento da compilação. Isso visa permitir que o desenvolvedor do programa original tenha a opção de exigir que seu código original não seja modificado. Nesse caso esses arquivos de correção podem ser distribuídos, e alterar o programa original somente no momento da tradução para o código de máquina entendido pelo computador. E o autor pode também exigir que o trabalho derivado seja distribuído sob uma nova nomenclatura ou sob um diferente número de versão.
* A licença não pode discriminar nenhum tipo de pessoa ou grupo.
* A licença não pode restringir qualquer tipo de uso do programa.
* Os direitos sobre o programa devem ser definidos somente através da licença, não sendo permitido nenhum mecanismo adicional. Isso visa evitar que o programa tenha uma licença livre, mas que o usuário seja obrigado, por exemplo, a concordar com um termo de não divulgação.
* A licença não pode ser condicional à presença do programa numa determinada distribuição ou conjunto de outros programas. Se o programa licenciado for retirado dessa distribuição, a licença deve permanecer válida mesmo que ele esteja atuando separadamente.
* A licença não deve exigir que outro programa distribuído em conjunto siga a mesma licença. Um programa de código aberto pode, portanto, ser distribuído em conjunto com um proprietário sem que este tenha que se tornar livre.
* A licença deve ser tecnologicamente neutra, não devendo exigir o uso de uma tecnologia específica ou tipo de interface (p.26,27).

## Implantação

Segundo Peppers & Rogers Group, o CRM é uma iniciativa corporativa. Ele deve ser planejado com cuidado para não haver falhas. Treinar adequadamente toda empresa é fundamental para o sucesso da implantação do CRM.

Implantar CRM na empresa começa com uma decisão estratégica de negócio.

Para Peppers & Rogers Group (2012), o primeiro passo para uma empresa construir uma sólida iniciativa orientada a clientes é ter uma visão e uma estratégia de clientes bem definidas.

Segundo o site s.a. builder (2012), o caminho mais seguro para implantar o CRM é envolvendo a direção principal, que tem a visão do todo e entende a importância de todas as informações de cada parte da empresa.

# QUADRO METODOLÓGICO

## Tipo de Pesquisa

Nosso projeto será de forma aplicada.

## Contexto de pesquisa

A implantação será realizada em uma empresa, que visa melhorar o relacionamento com os clientes.

## Participantes

Este projeto terá a participação de Aline Mara de Souza, Hellen de Souza Castro sobre orientação da professora Crishna Irion e com contribuição da responsável pela área a ser implantada o software.

## Produção de Dados

Reuniões quinzenais com a orientadora do projeto, com o objetivo de enfatizar pesquisas e dúvidas.

Reunião com o responsável da empresa, para a obtenção dos requisitos para a implantação.

## Procedimentos

Para que a implantação seja ocorra:

* A análise de requisitos.
* A pesquisa e escolha do CRM.
* O estudo e analise do software escolhido.
* Testes.
* Apresentação ao responsável da empresa.
* Implantação na empresa.
* Treinamento aos usuários.
* Assistência ao uso inicial do software.
* Análise final dos resultados após a implantação.

## Cronograma

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mês** | **FEV** | **MAR** | **ABR** | **MAIO** | **JUN** |
| **Tarefas** |  |
| Orientação do Pré-projeto | | x |  |  |  |  |
| Formulação do Pré-projeto | | x |  |  |  |  |
| Pesquisas dos itens do pré - projeto | | x |  |  |  |  |
| Fechamento do pré-projeto | |  | x |  |  |  |
| Entrega do pré-projeto | |  | x |  |  |  |
| Orientação da Introdução, Objetivos e Justificativa | |  | x |  |  |  |
| Formulação da Introdução, Objetivos e Justificativas | |  | x |  |  |  |
| Fechamento da Introdução, Objetivos e Justificativas | |  | x |  |  |  |
| Entrega da Introdução, Objetivos e Justificativas | |  |  | x |  |  |
| Orientação do Quadro Teórico | |  |  | x |  |  |
| Formulação do Quadro Teórico | |  |  | x |  |  |
| Entrega do Quadro Teórico | |  |  |  | x |  |
| Orientação do Quadro Metodológico | |  |  |  | x |  |
| Formulação do Quadro Metodológico | |  |  |  | x |  |
| Entrega do Quadro Metodológico | |  |  |  | x |  |
| Revisão do projeto para entrega | |  |  |  |  | x |
| Entrega dos projetos para qualificação | |  |  |  |  | x |
| Bancas de qualificação de Projetos | |  |  |  |  | x |
| Orientações finais dos projetos | |  |  |  |  | x |

# REFERÊNCIAS

SCUP IDEAS, O Voto Conectado. <http://scup.com/ideas/eleicoes-2014>, acessado em 05 de fevereiro de 2015.

CARTACAPITAL, Política. <http://www.cartacapital.com.br/politica/redes-sociais-ganham-importancia-mas-ainda-nao-definem-eleicao-5440.html>, acessado em 06 de fevereiro de 2015.

SILVEIRA, SERGIO AMADEU da, Software Livre: A luta pela liberdade do conhecimento, São Paulo, Editora Fundação Perseu Abramo, 2004.

SWIFT, Ronald, CRM, Customer Relationship Management: O Revolucionário Markenting de Relacionamento com o cliente, tradução de Flávio Deny Steffen,Rio de Janeiro, Elsevier 2001