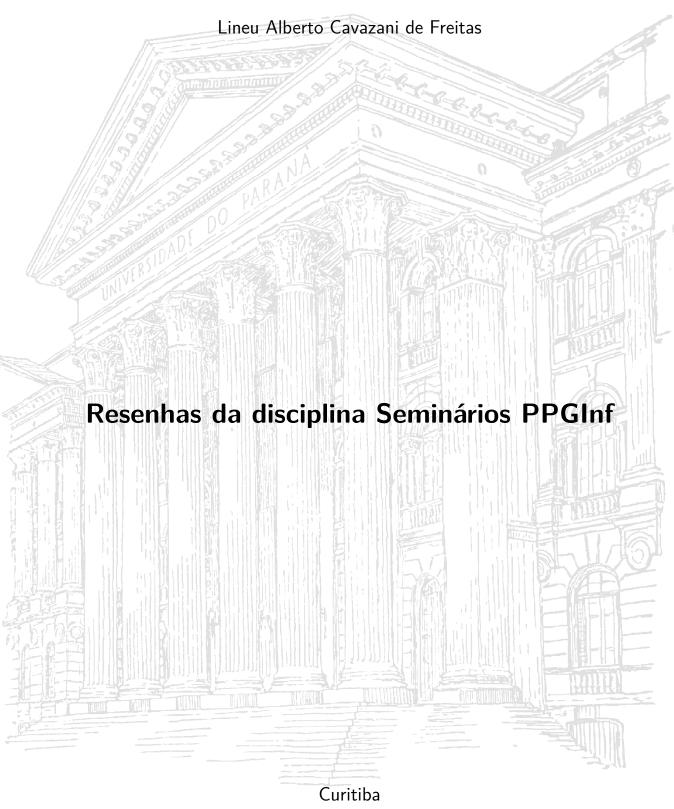
#### Universidade Federal do Paraná



2020

#### Lineu Alberto Cavazani de Freitas

#### Resenhas da disciplina Seminários PPGInf

Trabalho apresentado à disciplina de seminários do Programa de Pós Graduação em Informática da Universidade Federal do Paraná.

Universidade Federal do Paraná
Setor de Ciências Exatas
Programa de Pós graduação em Informática

Professor: Dr. Luis Carlos Erpen Bona

Curitiba 2020

#### Sumário

1	COVID-19 – CONCEITOS DE EPIDEMIOLOGIA 5
2	FORTES CORRELAÇÕES ENTRE O CRESCIMENTO (TIPO LEI DE POTÊNCIA) DA COVID-19 EM QUATRO CONTINENTES E A INEFICIÊNCIA DAS ESTRATÉGIAS SUAVES DE QUARENTENA 7
3	OVERVIEW DAS PESQUISAS MULTIDISCIPLINARES DO SETOR DE EXATAS EM ÉPOCA DE PANDEMIA 9
4	MULHERES NAS CIÊNCIAS: AÇÕES DE PESQUISA E EXTENSÃO 11
5	SOFTWARE MODERNIZATION
6	LIGADOS PELA LINGUAGEM: DIÁLOGOS POSSÍVEIS ENTRE LINGUISTAS E CIENTISTAS DE COMPUTAÇÃO 15
7	DANÇANDO COM ROBÔS: PLANEJAMENTO AUTOMÁTICO DE MITIGAÇÃO INTERMITENTE PARA COVID-19 17
8	VALIDAÇÃO DE SISTEMA DE APRENDIZADO DE MÁQUINAS PARA AUXÍLIO AO DIAGNÓSTICO DO COVID-19 BASEADO EM IMAGENS TORÁCICAS DE ARQUIVO RADIOLÓGICO 19
9	IMPACTO ECONÔMICO DA COVID-19 NO BRASIL: PODEMOS CULPAR O ISOLAMENTO SOCIAL?
10	CLASSIFICAÇÃO AUTOMÁTICA DE CÂNCER DE MAMA A PAR- TIR DE IMAGENS HISTOPATOLÓGICAS
11	GREENFIELD: PESQUISA CIENTÍFICA EM RECURSOS MINE- RAIS, ENERGIA E MEIO AMBIENTE
12	CIÊNCIA DE REDES EM CIÊNCIA DE DADOS: DESAFIOS 27

#### 1 COVID-19 – Conceitos de epidemiologia

Resenha do seminário ministrado no dia 25/05/2020 no Programa de Pós Graduação em Informática da UFPR por Bernardo Montesanti Machado de Almeida(médico infectologista da Unidade de Vigilância em Saúde do CHC/UFPR) e Wagner Hugo Bonat (professor do Departamento de Estatística da UFPR).

No seminário foi apresentado o conceito de taxa de reprodução, denotada por  $R_t$  que estima a disseminação de uma doença infectocontagiosa. Trata-se do número esperado de transmissões que devem ocorrer a partir de um caso. Dito isso, existem 3 possíveis resultados:  $R_t = 1$ , isto é, um infectado transmite para apenas uma pessoa, representa o cenário de equilíbrio endêmico.  $R_t < 1$ , trata-se da chamada supressão endêmica. E, por fim,  $R_t > 1$  representa o cenário de ciclo epidêmico, no qual um infectado transmite para mais de um indivíduo o que gera a tão falada curva exponencial de casos. O ideal é atingir o cenário de supressão pois este indica queda progressiva na incidência considerando que sejam mantidas as condições daquele momento.

Contudo vale ressaltar que os valores de *R* são sensíveis à dinâmica da doença e a fatores como localidade e clima. Como exemplo de ilustração, o doutor Bernardo citou o impacto do fator climático na pandemia de H1N1, em 2009, que fez com que a taxa de reprodução fosse superior a 2 no Paraná enquanto que a taxa era consideravelmente menor no restante dos estados brasileiros com clima mais quente.

Na atual pandemia do novo coronavírus, a taxa de reprodução varia entre 2 e 3, e é consideravelmente mais alta do que outras pandemias históricas como H1N1, Ebola e gripe espanhola.

Tratando mais a respeito do Corona vírus: a COVID-19 é uma doença causada pelo coronavírus SARS-CoV-2, descoberto em 31/12/19 após casos registrados na China. De acordo com a Organização Mundial de Saúde (OMS), cerca de 80% dos infectados podem ser assintomáticos e cerca de 20% podem requerer atendimento hospitalar por apresentarem dificuldade respiratória.

Os sintomas da COVID-19 podem variar de um simples resfriado até uma pneumonia severa. A transmissão acontece por meio de toque (na pessoa ou em superfície contaminada), gotículas de saliva, espirro, tosse, catarro, etc. O diagnóstico é feito através de exames laboratoriais. Na prática, trata-se de um vírus com letalidade baixa mas altamente contagioso e que, até o momento, não há vacina. Portanto, caso a disseminação não seja devidamente controlada, é uma doença capaz de levar sistemas de saúde inteiros ao colapso como ocorreu em diversos países do mundo. Todavia, devido à variedade de sintomas, controlar a cadeia de transmissão se torna uma tarefa

difícil considerando que boa parte dos casos são assintomáticos, diferente de doenças conhecidas como Ebola, por exemplo, em que os sintomas são tão mais violentos que se torna uma tarefa fácil identificar o indivíduo contaminado e isolá-lo, acabando com a cadeia de transmissão que seria gerada por esse indivíduo.

Por esta razão justifica-se a estratégia de distanciamento social para conter a disseminação da doença, pela dificuldade de isolar apenas aqueles infectados dado que muitos casos são silenciosos e a taxa de reprodução é alta.

Por mais que em termos percentuais não seja uma doença com alto índice de casos graves e letalidade, a taxa de reprodução é a maior preocupação pois 20% de um número altíssimo de contaminados que terão sintomas graves, necessitarão de atendimento intensivo congestionando o sistema de saúde de tal forma que não haja leitos o suficiente para lidar com os casos graves de coronavírus e também para casos de outras doenças e circunstâncias que levem o indivíduo a necessitar de terapia intensiva; deste modo fazendo que a letalidade de todas as doenças que necessitam terapia intensiva aumente, e não apenas dos casos de COVID-19.

Outro fato que gera preocupação é o efeito que os cenários divulgados causam na população ao observar-se valores de taxas de contágio e resultados de modelos que apontam para o cenário de situação sob controle, pois estes resultados geram um afrouxamento das medidas de isolamento (atualmente nossa única medida protetiva na ausência de uma vacina) e causam a elevação das contaminações. Deste modo criase um ciclo vicioso, no qual as análises mostram que a situação dá indícios de estar sob controle, afrouxam-se as medidas restritivas, como resultado do afrouxamento a contaminação aumenta, o que nos obriga a retomar as medidas de isolamento e assim por diante.

Outro assunto que tem sido particularmente tratado é o termo imunidade de rebanho. A ideia consiste em imunizar boa parte da população de modo que os que não foram imunizados estejam protegidos. O problema é que com a ausência de uma vacina, este tipo de imunização se daria pelo contágio (a pessoa contaminada cria anticorpos e não contamina mais, está imune). Assim, imaginemos um cenário hipotético: digamos que para atingir imunidade de rebanho precisaríamos de, sendo otimista, 2/3 da população imunizada. Considerando que a população brasileira atualmente é aproximadamente 210 milhões, necessitaríamos de 138 milhões de pessoas (2/3) contaminadas. Contudo, destas 138 milhões, por volta de 27 milhões de pessoas necessitariam de atendimento intensivo devido à sintomas graves (20% dos infectados, como indica a OMS). Atualmente a letalidade no Brasil relacionada ao coronavírus está em torno de 6%, deste modo, considerando este cenário, haveriam 8 milhões de óbitos. Nota-se que é um cenário completamente inviável.

# 2 Fortes correlações entre o crescimento (tipo Lei de Potência) da COVID-19 em quatro continentes e a ineficiência das estratégias suaves de quarentena

Resenha do seminário ministrado no dia 28/05/2020 no Programa de Pós Graduação em Física da UFPR por Marcus Beims (professor do Departamento de Física da UFPR).

No seminário ministrado pelo professor Marcus Beims foi apresentado um breve contexto a respeito de modelos epidemiológicos, como o modelo de suscetíveis e imunes (SI), o modelo de suscetíveis, infectados e recuperados (SIR) e o modelo Suscetíveis, Expostos, Infectados e Recuperados (SEIR).

No trabalho foi utilizado o modelo SEIR para descrever a contaminação e estudar melhores estratégias para atingir o famoso achatamento da curva. Uma das conclusões é que modelos epidemiológicos podem ser descritos por redes de "mundo pequeno" ou "livres de escala".

Além disso foram apresentadas técnicas para obtenção de medidas de correlação entre tempo e número de casos acumulados e também do potencial de curvas de crescimento de casos de Covid-19 ajustadas via Lei de Potência ao invés da famosa curva exponencial, amplamente utilizada e atualmente discutida no atual cenário de pandemia.

Os resultados mostram que as estratégias de contenção são igualmente válidas em diferentes países o que refuta argumentos muito ditos no cenário atual de que deve-se regionalizar a política de isolamento no Brasil dado seu território.

Outro importante resultado ressalta a importância do isolamento social para contenção do alastramento da doença considerando o potencial de disseminação do vírus e a dificuldade de identificar e isolar infectados durante a fase de contaminação fazendo com que o isolamento seja, no cenário atual, nossa única linha de defesa na ausência de uma vacina.

Segundo o estudo, um distanciamento social mediano, junto com a identificação e isolamento diário de poucos dos indivíduos infectados assintomáticos, não é suficiente para diminuir significativamente o avanço do número de casos. Mostrando que as

demandas e desafios atuais são encontrar meios eficientes para se identificar indivíduos contaminados que sejam assintomáticos visando o isolamento destes para que não contaminem demais indivíduos e, enquanto isso não seja possível, aumentar e reforçar de forma significativa as medidas de distanciamento social.

Deste modo, dadas as circunstâncias e dificuldades no cenário atual, nossa melhor chance para diminuição do número de infectados são implementar eficazes medidas de isolamento social combinadas à realização periódica massiva de testes que viabilizam diagnosticar indivíduos infectados assintomáticos com o objetivo de isolá-los de forma efetiva de tal modo que a taxa de reprodução deste indivíduo seja neutralizada. Deste modo, a longo prazo a taxa de reprodução geral será reduzida e sairemos do risco de colapsar o sistema de saúde.

## 3 Overview das pesquisas multidisciplinares do Setor de Exatas em época de pandemia

Resenha do seminário ministrado no dia 29/05/2020 no Programa de Pós Graduação em Informática da UFPR por Marcos Sunye (diretor do setor de ciências exatas da UFPR).

No seminário ministrado pelo professor Marcos Sunye foram apresentadas as diversas iniciativas de cursos, programas de pós graduação, professores e pesquisadores das exatas em estudos e iniciativas com o objetivo de compreender, conter e divulgar informações de forma correta a respeito do novo corona vírus. O Setor de Ciências Exatas possui 6 departamentos, 7 programas de pós graduação e todos estão envolvidos de algum modo em trabalhos a respeito do Covid-19.

As ações vão desde a produção de álcool em gel feita pelo Departamento de Química, passando pelo uso de impressoras 3D pela Expressão Gráfica para impressão de materiais de segurança até estudos sobre a disseminação da doença e estratégias de contenção e diagnóstico.

Existe a iniciativa Portal Covid-19 que tem por objetivo agregar informações, modelos e visualizações de dados sobre a pandemia. Com contribuições de pesquisadores dos Departamentos de Estatística, Informática, Física, Matemática e Saúde da Universidade.

Além disso existem projetos vinculados à pesquisadores da Física como o do professor Marcus Beims que faz uso do modelo SEIR para entender de que forma se propaga o vírus e sugerir medidas e estratégias para conter sua disseminação.

Há também projetos com foco em testagem, como o do professor David Menotti que faz uso de técnicas de processamento de imagem, visão computacional e aprendizado de máquinas para diagnóstico em que a entrada é uma radiografia do pulmão do paciente.

E diversos outros projetos em andamento que avaliam a viabilidade e como proceder de forma segura e eficiente o retorno das atividades, como de que forma manter a disposição dos laboratórios e técnicas para aferição de temperatura dos alunos que entram nas instalações da Universidade, por exemplo.

### 4 Mulheres nas Ciências: ações de Pesquisa e Extensão

Resenha do seminário ministrado no dia 01/06/2020 no Programa de Pós Graduação em Informática da UFPR por Camila Silveira da Silva (professora do Departamento de Química da UFPR).

No seminário mistrado pela professora Camila Silveira foram abordados aspectos da evidente desigualdade de gênero nas ciências de modo geral e, mais especificamente, nas ciências exatas, onde a ausência de equidade é ainda mais gritante.

Numa primeira provocação a professora sugeriu que fizéssemos um questionamento: quantas mulheres cientistas você conhce? O que prontamente nos mostra como a imagem de cientista está vinculada à figura esteriotipada de homem branco no laboratório.

Após este exercício foram apresentadas uma série de motivos da necessacidade de se falar sobre mulheres na ciência na atualidade como o fato de apenas 30% de pesquisadores de ciência e tecnologia serem do sexo feminino, percentual bastante distante de um nível real de representatividade. Além disso, desde 1901 apenas 5% dos prêmios Nobel foram destinados a pesquisadoras.

Foi discutida também a ausência políticas que busquem igualdade de gênero nas universidades e a falta de incentivo ao sexo feminino para se fazer ciência, o que vem desde a infância com brinquedos temáticos sendo destinados à seção masculina de lojas infantis e vão até os desafios enfrentados por mulheres na academia, como comentários inadequados feitos por colegas homens e maternidade, grande fator de evasão entre as mulheres.

Dentre os 17 objetivos listados pela ONU visando desenvolvimento sustentável até 2030, um deles diz respeito à igualdade de gênero e tem como metas acabar com a desigualdade de gênero, eliminar toda e qualquer forma de violência contra a mulher, garantir a participação de mulheres com igualdade de oportunidade em todas as esferas da sociedade desde recursos econômicos até oportunidade de liderança na vida política, econômica e pública.

Por fim, foram apresentados uma série de movimentos atuais que buscam atingir estes objetivos nacional e internacionalmente com ênfase no Projeto de Extensão da Universidade Federal do Paraná Meninas e Mulheres na Ciência que conta com a participação de professoras e docentes de graduação e pós graduação das mais diversas áreas da Universidade.

#### 5 Software Modernization

Resenha do seminário ministrado no dia 01/06/2020 no Programa de Pós Graduação em Ciência da Computação da UTFPR por Wesley Klewerton (professor na UTFPR).

O trabalho apresentado pelo professor Wesley Klewerton tinha como objetivo definir e implementar uma metodologia automatizada para re-engenharia de um conjunto de modelos dentro de um Software Product Line Engineering (SPL).

Engenharia de Software área da Computação que se propõe a estudar e propor soluções para desenvolver sistemas de software de forma produtiva. Tarefa que em geral envolve partes complexas como produto (o que será o resultado final), processo (como fazer) e autores (individual, equipe, etc). Além disso, na Engenharia de Software, deve haver uma preocupação com a qualidade interna (aquela que diz respeito aos desenvolvedores) e externa (que diz respeito ao usuário).

Software modernization é uma área de Engenharia de Software que trata de problemas como conversão, reescrita ou portabilidade de sistemas para uma versão mais moderna de linguagem, protocolos ou hardware. Pode se tratar de uma tarefa custosa devido ao risco, dificuldade e à ausência de uma ferramenta de propósito geral para realização desta tarefa (por mais que hajam algumas técnicas específicar para fazê-lo).

A necessidade da modernização pode levar à re-engenharia, migração, substituição e reuso. A reutilização trata-se da reciclagem, isto é, aproveitar um projeto já existente e modificá-lo para obter outro produto final. Existem técnicas como Software Product Line Engineering (SPL) que é um paradigma para sistematização das técnicas de reuso. Já a re-engenharia é a análise e alteração de um sistema a fim de sua reestruturação em uma nova forma e, este trabalho para SPLs consiste de diferentes fases e propósitos. Trata-se de um tema complexo e de literatura limitada e , no trabalho, foi proposta uma implementação de metodologia automatizada para re-engenharia de um conjunto de modelos dentro de um SPL.

## 6 Ligados pela linguagem: diálogos possíveis entre linguistas e cientistas de computação

Resenha do seminário ministrado no dia 05/06/2020 no Programa de Pós Graduação em Informática da UFPR por Adelaide H. Pescatori Silva (professora do Departamento de Letras UFPR e pós-doc PPGInf).

No seminário ministrado pela professora Adelaide foram tratados num primeiro momento as similaridades entre linguística (estudo da linguagem humana) e computação (análise de dados através de linguagens artificiais) e a sua relação com um mesmo objeto de interesse: a linguagem.

Foi apresentado que a linguística é construída por partes ou módulos e a compreensão da estrutura geral e de cada módulo auxilia na compreensão do todo, enquanto que, de modo similar, na programação, apenas alguns pequenos blocos de funcionalidades, quando combinados, são suficientes para solucionar diversos tipos de problema. Outro ponto interessante comentado foi a semelhança entre os diagramas do processamento da fala humana por um falante e o processamento de dados feito por um computador e que, em ambos os casos, há um papel fundamental da memória.

Um ponto histórico em que a convergência entre linguística e computação resultou num produto foi o desenvolvimento de "máquinas falantes" que, dentre as possíveis funcionalidades pode-se citar síntese, tradução e até mesmo a própria comunicação entre humano e máquina. Além de outros produtos gerados ao longo do tempo como ferramentas de tradução automática, reconhecimento e interpretação de fala, classificação e produção de textos, sistemas de busca, etc.

Foi apresentada também a grande similaridade entre o processo de estudos a respeito de linguística e o processo percorrido pela ciência de dados, que consiste em: fazer uma pergunta (hipótese), coletar dados, tratá-los e explorálos, buscar modelos que permitam inferir os resultados amostrais para uma população e por fim, interpretar os resultados de modo a verificar se a hipótese inicial foi ou não aceita, isto é, se há evidência suficiente para rejeitar ou não a hipótese inicialmente levantada.

Por fim a professora Adelaide apresentou o plano do seu trabalho de pós-doc: a criação de um jogo para fixação de normas ortográficas de modo contextualizado em que o público alvo seriam alunos do Fundamental II (6° ao 9° ano) que faz uso de linguística probabilística e técnicas de aprendizado de máquina. Um plano altamente motivado pelo atual cenário de pandemia em que se fazem necessárias ferramentas educacionais remotas.

## 7 Dançando com robôs: planejamento automático de mitigação intermitente para Covid-19

Resenha do seminário ministrado no dia 05/06/2020 no Programa de Pós Graduação em Matemática da UFPR por Paulo J.S. Silva (professor do IMECC-Unicamp).

No seminário o professor Paulo foram apresentados estratégias e modelos que visam lidar de forma mais eficiente com questões como isolamento social em tempos de pandemia de tal modo que proteja a população do contágio e sem que a economia seja tão afetada.

A discussão iniciou-se na contextualização dos protocolos de distanciamento social incialmente aplicados: tratam-se de protocolos que adiam o problema, considerando que não geram imunização da população. Além disso os protocolos iniciais eram ineficientes por não analisar questões como a infraestrutura de cada regionalidade. No estado de São Paulo foi implantado um novo protocolo que olha para a capacidade de cada local. Deste modo, o objetivo era chegar a um modelo que pudesse organizar um isolamento social de forma regionalizada, conforme a capacidade de atender casos de contaminação e deste modo mantendo parte da economia constantemente ativa e evitando que toda a sociedade mantenha-se isolada ao mesmo tempo.

Após este contexto foi apresentado um pouco a respeito dos modelos compartimentais, iniciando pelo modelo SIR em que é necessária a quantidade de indivíduos suscetíveis, infectados e recuperados. Através destas quantidades é possível obter também a taxa de reprodução basal ( $R_0$ ) que representa o número médio de pessoas que um infectado em fase de contaminação pode afetar. Além do modelo SIR, foi apresentado o SEIR, no qual acrescenta-se um compartimento referente aos expostos, sendo assim existem os compartimentos suscetíveis, expostos, infectados e recuperados.

A proposta de modelo apresentada é uma variação do modelo SEIR que considera o deslocamento de indivíduos entre municípios de tal modo que as quantidades referentes aos compartimentos sejam alteradas durante o dia e, durante a noite, considera-se que os indivíduos voltam para seu município de residência. Com o modelo proposto é possível obter valores de  $R_t$  para verificar o número de "vidas salvas" graças às políticas de isolamento implantadas em determinado período de tempo e estudar perfis de mitigação da disseminação da doença avaliando diferentes cenários de distanciamento regionalizado.

Por fim, foi apresentado um estudo de caso em que se fez uso o modelo proposto. Foram avaliadas 11 cidades fortemente relacionadas e com altos índices de deslocamento entre as cidades (São Paulo e região metropolitana) em que  $R_0$  estava próximo de 1.5. Uma avaliação de valores de R sugerem que valores próximos a 2 mostram um isolamento social baixo, 1.4 elevado até chegarmos no 1 que é visto como isolamento severo.

Utilizando o modelo SEIR modificado foi possível lidar com o deslocamento entre os municípios, avaliar o cenário atual e diversos possíveis cenários de políticas de distanciamento social adaptativas que levariam a uma forma eficiente de utilização dos recursos de tal modo que o sistema seja capaz de lidar com os contaminados nesses pequenos períodos tempos sem sobrecarga.

#### 8 Validação de Sistema de Aprendizado de Máquinas para auxílio ao diagnóstico do COVID-19 baseado em imagens torácicas de arquivo radiológico

Resenha do seminário ministrado no dia 08/06/2020 no Programa de Pós Graduação em Informática da UFPR por David Menotti (professor do Departamento de Informática UFPR).

No seminário ministrado pelo professor David foram tratados aspectos sobre o Inteligêcia Aritificial, Machine Learning e Deep Learning. Entende-se por Inteligência Artificial o ramo que busca mecanismos e dispositivos com o objetivo de simular a capacidade do ser humano de raciocinar e resolver problemas. Dentro da Inteligência Artificial existe o Aprendizado de Máquina (Machine Learning) em que um modelo analítico é treinado para prever ou classificar resultados futuros.

Um tipo específico de Machine Learning é chamado de Deep Learning que tem como objetivo treinar máquinas para que reconheçam padrões de forma mais próxima a como um ser humano faria, como por exemplo: reconhecimento de fala, identificação de imagem e previsões.

Máquinas não tem consciência, contudo são capazes de aprender e isso é mostrado por dispositivos e produtos que surgiram ao longo da história como veículos autônomos.

Na prática, tanto máquinas quanto seres humanos aprendem da mesma forma: a partir de bons exemplos e correção em caso de erro. Como ilustração de possíveis métodos apresentados para esse fim, no contexto de aprendizado de máquina, foram apresentadas as Redes Neurais Artificiais e os Support Vector Machines.

Já no contexto de Deep Learning fatores que influenciam para adequada execução são o desenvolvimento de algoritmos, capacidade de hardware para processar grandes massas de dados e disponibilidade de dados adequados. A grande diferença entre as técnicas de Machine Learning tradicionais e o Deep Learning é a característica dos descritores. Nas técnicas tradicionais os descritores são fixos, enquanto que nas técnicas de aprendizado profundo são treináveis.

Após esta contextualização foram apresentados diversos eventos importantes no

que diz respeito às técnicas de Deep Learning. Por fim, o professor David falou a respeito do trabalho desenvolvido em tempos de pandemia cujo objetivo era desenvolver um sistema de aprendizado de máquinas para auxílio ao diagnóstico do COVID-19 baseado em imagens torácicas de arquivo radiológico e as dificuldades de continuidade do trabalho devido ao impasse com comitês de ética.

#### 9 Impacto econômico da COVID-19 no Brasil: podemos culpar o isolamento social?

Resenha do seminário ministrado no dia 19/06/2020 no Programa de Pós Graduação em Informática da UFPR por Luis Celso Gomes (professor do Departamento de Informática UTFPR).

No seminário ministrado pelo professor Luis Celso foram abordados aspectos econômicos da pandemina de Covid-19, uma discussão motivada pelo fato de que o Brasil vem se baseando na postura de países mais ricos que passaram pela pandemia. O problema desta abordagem é que as condições financeiras destes países permitem uma recuperação menos demorada, o que não é possível no Brasil.

Como disse o professor, trata-se de uma crise que não é comparável com nenhuma outra da história pois nunca houve uma crise tão profunda, que afetasse o mundo por completo e ainda associada a um problema de saúde com fim incerto.

Tomando os Estados Unidos como exemplo, trata-se do maior índice de desemprego causado por uma crise. O ponto é que se um país rico entra em recessão, um cidadão médio retorna à sua antiga condição financeira de forma mais rápida que o cidadão médio de um país em desenvolvimento, como o Brasil.

Mas até que ponto podemos culpar as medidas restritivas pela condição econômica do país? Indicadores mostram que o Brasil é um dos países mais afetados pela pandemia e sofreu impacto econômico maior do que países do BRICS e sofre mais que a média mundial. Porém vale ressaltar que antes do vírus chegar ao Brasil já podiam ser observados efeitos na economia causado pelo impacto na pandemia nos países desenvolvidos.

Outro fator importante relacionado ao impacto da pandemia é a implementação de medidas menos pesadas de isolamento social. Quando comparados os índices de redução de poluição causada por automóveis, é possível notar que o Brasil possui uma das menores reduções no início da pandemia. Mostrando que quando o vírus chegou, poucas pessoas se mantiveram em isolamento.

Logicamente, a ausência de medidas mais efetivas para controle da doença foi refletida no cenário que vemos hoje, em que o Brasil se tornou o epicentro da doença no mundo e líder em número diário de óbitos.

Um resultado interessante é a relação entre redução de mobilidade e o avanço no número de óbitos que mostra que a medida que reduz-se a circulação de indivíduos reduz-se também o número de óbitos causados pela doença, mostrando que medidas de isolamento são altamente necessárias para preservar vidas.

Por fim, o professor Luis mostrou um estudo de caso a respeito do setor de aviação. Trata-se de um setor altamente rentável e que, mundialmente movimenta mais capital que países inteiros. Contudo foi um dos setores mais afetados com a pandemia a nível mundial devido à menor circulação de indivíduos, seja dentro do país ou para fora dele. O ponto central da discussão a respeito do impacto da aviação foi de que no começo da pandemia fazia sentido restringir a circulação de indivíduos a nível mundial com o intuito de impedir a disseminação, isto é, não permitir que pessoas contaminadas cheguem a diferentes lugares. Contudo, à medida que o globo é afetado por inteiro esta medida perde importância pois chega-se a um nível que a doença já se espalhou e não faz mais sentido reduzir o tráfego aéreo para fins de redução de contágio.

## 10 Classificação automática de câncer de mama a partir de imagens histopatológicas

Resenha do seminário ministrado no dia 22/06/2020 no Programa de Pós Graduação em Ciência da Computação da Unioeste por Fábio Spanhol.

No seminário ministrado pelo professor Fábio Spanhol foi apresentado o conteúdo referente ao sistema de classificação de câncer de mama desenvolvido no seu doutorado na UFPR.

Dentro da área de reconhecimento de padrões em exames médicos de imagem para fins de diagnóstico, destacam-se as iniciativas para identificação de casos de câncer por se tratar de uma doença extremamente comum e de alta letalidade.

Câncer é uma nomenclatura geral para designar doenças caracterizadas pela replicação celular anormal que gera tumores. Estes tumores podem ser benignos ou malignos, os malignos recebem o nome de câncer.

Sabe-se que em países desenvolvidos predominam tipos de câncer ligados à urbanização, como por exemplo o câncer de pulmão, próstata, mama, etc. Já em países menos desenvolvidos os casos mais comuns são de tipos de câncer ocasionados por falta de recursos como saneamento básico: estômago, fígado, etc.

No Brasil, a maior parte dos casos de câncer entre homens é na próstata enquanto que nas mulheres o câncer de mama é o mais comum. O trabalho do professor Fábio refere-se ao câncer de mama; trata-se de um tumor maligno que afeta o tecido mamário em que se destacam o carcinoma ductual (80% dos casos) e o lobular (entre 10% e 15% dos casos). A detecção se dá por meio de métodos não invasivos (diagnóstico por imagem) ou invasivos (biópsia).

A pesquisa utiliza dados referentes a biópsias do tipo cirúrgica na qual um médico patologista analisa a amostra a fim de concluir se trata-se de um tumor maligno ou benigno. Dito isso, o objetivo do trabalho era desenvolver uma ferramenta de auxílio ao patologista para realizar o diagnóstico de uma amostra coletada via biópsia cirúrgica de um caso suspeito de câncer de mama.

Um primeiro produto deste trabalho foi a geração de uma base de dados pública com imagens histopatológicas de tumores mamários tanto malignos quanto benignos: a BREAKHIS. Atualmente mais de 1000 usuários utilizam estes dados em todo o mundo.

O segundo produto do trabalho foi o desenvolvimento de um sistema de classificação de tipo de câncer de mama. Dentre as técnicas conhecidas de Machine Learning destacou-se o support vector machine, com acurácia em torno de 85%. Utilizando técnicas de deep learning, mais especificamente com redes neurais convolucionais, houve resultados de acurácia acima de 90%.

## 11 Greenfield: Pesquisa científica em recursos minerais, energia e meio ambiente

Resenha do seminário ministrado no dia 26/06/2020 no Programa de Pós Graduação em Informática da UFPR por Leonardo Cury (professor do Departamento de Geologia/LAMIR UFPR).

No seminário ministrado pelo professor Leonardo foi apresentado o conceito de Exploração Greenfield. Trata-se da exploração de recursos naturais em áreas inexploradas, de forma sustentável, ao contrário de um imaginário popular que crê tratar-se e uma exploração desenfreada e mal planejada dos rescursos disponíveis.

Na primeira parte parte da apresentação foram tratados aspectos práticos, isto é, a ida ao campo, na qual coletam-se dados, amostras e avalia-se pessoalmente o ambiente de interesse a fim de obter material suficiente para observação em diferentes magnitudes, o que permite extrapolar os resultados para uma área maior. Na prática, um projeto parte de uma ideia promissora. Quanto mais promissora a ideia, maior a chance de investimento. Como exemplo o professor Leonardo citou o caso Brasil x Angola, em que a hipótese é: se existe óleo e gás do lado brasileiro, será que também existe do lado angolano? Considerando que um dia tratou-se de uma mesma região, é possível. Contudo cabe explorar, pois por mais que sejam unidades cronocorrelatas existem diferenças como, por exemplo, as esperruras do pré-sal nas duas regiões.

Na segunda parte foram tratados aspectos pré-campo, que consistem em conhecer melhor o terreno através de técnicas a fim de chegar a mapas temáticos geológicos. Neste contexto, mapas são mais que desenhos, são conjuntos de dados e por isso devem ser muito bem planejados e passar por etapas como: a concepção de um tema, a implementação do raciocínio, a obtenção e análise dos dados, a extração dos temas e a correção cartográfica.

Além do trabalho de campo é de igual importância o trabalho de laboratório. Este exige grande infraestrutura com aparelhos caros e, certas vezes, de difícil uso no sentido de automação de rotinas. Contudo, esta estrutura gera uma quantidade imensa de dados e o planejamento de como organizá-los é essencial.

Na parte da apresentação destinada às áreas de pesquisa foi mostrada uma série de exemplos de locais interessantes com propriedades bastante especiais com potencial para desenvlvimento de trabalhos de pesquisa promissores. E por fim uma breve discussão sobre aplicação no contexto de dados de subsuperfície.

#### 12 Ciência de Redes em Ciência de Dados: Desafios

Resenha do seminário ministrado no dia 09/07/2020 no CiDAMO (grupo de Ciência de Dados, Aprendizagem de Máquina e Otimização da UFPR) por Artur Ziviani (professor do Laboratório Nacional de Computação Científica - LNCC).

O seminário do professor Artur iniciou-se com uma apresentação a respeito do Laboratório Nacional de Computação Científica, o LNCC. Trata-se de uma unidade de pesquisa do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC) localizada em Petrópolis, no estado do Rio de Janeiro. Conta atualmente com aproximadamente 50 pesquisadores das mais diversas formações, tais como: engenharias, ciência da computação, matemática aplicada, física, biologia, etc.

Considerando o número de pesquisadores filiados, o LNCC é uma unidade pequena mas que possui pós graduação em modelagem computacional com mestrado e doutorado. Além disso possui conceito CAPES 6, conta com diversos pesquisadores em pós-doutorado, além de muitos trabalhos de pesquisa e desenvolvimento em cooperação com instituições nacionais e internacionais. No LNCC encontra-se a infraestrutura computacional mais poderosa da América Latina destinada a fins acadêmicos.

No que diz respeito à área de pesquisa de modelagem computacional, dentre as diversas vertentes possíveis encontra-se a Ciência de Dados. Para trabalhos nesta área em especial existe no LNCC o DEXL LAB (Data Extreme Lab), gerenciado pelos professores Fabio Porto, Artur Ziviani e Luiz Gadelha, no qual os 3 principais interesses são Big Data, Inteligência Artificial e Modelagem Matemática e Computacional. Os tópicos de pesquisa abrangem temas como: gerenciamento de Big Data, modelagem de dados, algoritmos para análise de dados, ciência de redes, aprendizado de máquina, etc. O laboratório de Ciência de Dados do LNCC contribui em projetos de áreas como: astronomia, saúde pública, ciência do esporte, petróleo e gás, etc.

No que diz respeito aos desafios encontrados dada a popularização e geração de grande quantidade de dados, o professor Artur citou que necessitamos de formas eficientes de lidar com os 4 Vs: Volume, Velocidade, Variedade e Veracidade. Como exemplo de fontes das quais surgem dados foi dado o exemplo de redes sociais, nas quais a todo momento são coletadas informações dos usuários para os mais diversos fins, ou seja, nós somos uma enorme fonte de informação.

Por fim foi apresentada a ideia geral de que que redes complexas correspondem a um grafo, que é representado por um conjunto de nós ligados por arestas, que em conjunto representam relações. Após o conceito foram mostrados diversos estudos desenvolvidos no LNCC que fazem uso de tal metodologia, como por exemplo: aplicação de redes complexas a um problema de biologia, uso de técnicas de redes complexas para melhoria de algoritmos de aprendizado de máquina, etc.