Desenvolvimento de um ETL para treinamento de modelos QSAR a partir do PubChem BioAssays

JOÃO PEDRO GOMES GRECO[1]; FREDERICO SCHMITT KREMER[2]

[1]*Laboratório de Bioinformática, Graduação em Biotecnologia, CDTec, UFPel – joao.greco@ufpel.edu.br*
[2]*Laboratório de Bioinformática, Graduação em Biotecnologia, CDTec, UFPel – fred.s.kremer@gmail.com*

1. DESCRIÇÃO DA INOVAÇÃO

A descoberta de novos medicamentos, ou *Drug Discovery Pipeline* (DDP) é essencial para o avanço da medicina, mas é um processo complexo devido à necessidade de triagem de bilhões de compostos (Attene-Ramos, Austin, Xia, 2014). Para acelerar este processo, métodos como relação quantitativa estrutura-atividade (QSAR) e modelagem molecular, ou docking têm se mostrado eficazes. QSAR prevê a atividade biológica de compostos químicos, enquanto o docking analisa a interação entre moléculas e seus alvos biológicos, ambos economizando tempo e recursos(Neves et al., 2018).

Bancos de dados científicos como PubChem (https://pubchem.ncbi.nlm.nih.gov/) e o ChEBML (https://www.ebi.ac.uk/chembl/) armazenam enormes quantidades de dados de informações químicas e biológicas. Devido a esses grandes volumes armazenados, a automação auxilia de forma eficaz no gerenciamento dessas informações.

Nesse contexto, processos como o ETL (Extract, Transform, Load) são cruciais, pois realizam extrações de dados de diversas fontes., transformá-los para garantir a consistência e qualidade e carregar para sistemas de análise. Essa automação facilita a integração e a atualização contínua de dados complexos (Souibgui et al., 2019).

Um dos maiores desafios na descoberta de fármacos é a organização e integração de grandes volumes de dados. A demora na obtenção de dados pode atrasar significativamente o processo de descoberta. Para enfrentar esses desafios, métodos de engenharia de dados para ETL podem ser utilizados. O ETL desempenha um papel crucial ao combinar dados de várias fontes do sistema de saúde, padronizando os dados brutos em um formato uniforme e aplicando regras predefinidas para filtragem, padronização e normalização dos dados, visando armazenamento e análise futura.

Dentro desse contexto da descoberta de novos medicamentos, técnicas de ETL permitem automatizar a extração, transformação e carregamento de dados. Ao implementar o ETL, é possível integrar e padronizar dados provenientes de diversas fontes, facilitando a análise e o desenvolvimento de modelos QSAR. Dessa forma, o ETL não apenas acelera o processo de descoberta de fármacos, mas também garante a qualidade e a consistência dos dados utilizados, permitindo que os pesquisadores se concentrem nas etapas críticas de triagem e validação dos compostos.

2. ANÁLISE DE MERCADO

O crescimento do mercado de descoberta de medicamentos está em constante crescimento devido à demanda cada vez maior por novos tratamentos e avanços tecnológicos cada vez mais constantes. As empresas farmacêuticas como Johnson & Johnson, Novartis e Novartis (https://forbes.com.br/listas/2015/07/15-maiores-empresas-farmaceuticas-do-mundo/) estão na liderança da pesquisa e desenvolvimento de novos medicamentos, utilizando técnicas como ETL, pequenas empresas de biotecnologia também desempenham um papel crucial na utilização dessas técnicas para o desenvolvimento dos seus produtos.

A plataforma Bambu Enterprise é focada para atender uma ampla gama de pesquisadores, empresas farmacêuticas, instituições acadêmicas e laboratórios de biotecnologia que atuam na descoberta de novos fármacos através da utilização de ferramentas de bioinformática. Devido sua interface gráfica altamente simplificada e o uso de ferramentas colaborativas, o Bambu Enterprise se destaca tanto para auxiliar grandes empresas, pequenas empresas e laboratórios universitários que atuam na busca de inovação com baixo custo, alta eficiência e alternativas mais eficazes.

Empresas de bioinformática focadas para o desenvolvimento de softwares de bioinformática como Schröndiger, Certara e ChemAxon, que oferecem plataformas mais complexas de triagem de moléculas e descobertas de medicamentos para usuários leigos na área de bioinformática. Porém essas empresas exigem o uso de licenças para o uso dos seus produtos que acabam muitas vezes sendo bastante onerosos para microempreendedores. O Bambu Enterprise disponibiliza para universitários e instituições acadêmicas uma licença sem custos, além de possuir uma interface mais simples para o uso de usuários novos de bioinformática, deixando a ferramenta mais acessível para equipes que visam trabalhar e aprender a área de desenvolvimento de novos fármacos.

O mercado global de descoberta de medicamentos é estimado em cerca de US$100,10 bilhões em 2024, com expectativa de atingir US$137,72 bilhões até 2029. A taxa de crescimento anual composta (CAGR) prevista para o período de 2024 a 2029 é de 6,59% (https://www.mordorintelligence.com/pt/industry-reports/drug-discovery-market).

Esse aumento é cada vez mais impulsionado por investimentos farmacêuticos e a utilização de serviços terceirizados. O mercado pode ser dividido em segmentos que se encontram os fármacos biológicos, fármacos específicos, genéricos, similares e novos. Devido a esse aumento da demanda de novos fármacos, as perspectivas, na nossa plataforma apresenta um grande potencial para podermos preencher o espaço de mercado para auxiliar as necessidades dos pesquisadores e empresas.

3. ESTRATÉGIA DE DESENVOLVIMENTO E IMPLEMENTAÇÃO

O Bambu Enterprise utiliza um modelo de negócios baseado em oferecer acesso gratuito para pesquisadores e startups e para grandes empresas

farmacêuticas que atuam na vanguarda do mercado terão planos pagos com funcionalidades mais avançadas, para um maior processamento de dados e suporte técnico para o uso da ferramenta. O foco da distribuição da plataforma é online, sem a necessidade de instalação para tornar seu acesso mais acessível e parcerias com instituições universitárias.

A ferramenta já possui seu software registrado, com a permissão prevista para ser disponibilizada em breve, o que garante a proteção intelectual. Hoje em dia se encontra com a TLR 5, sendo seus protótipos funcionais já testados, a expansão e a melhora da ferramenta serão baseadas no feedback que os usuários irão fornecer ao utilizarem a ferramenta, focando em atingir o TRL 8. Os maiores desafios que estão sendo enfrentados vem sendo a aceitação por grandes empresas, escalabilidade da plataforma, que serão amenizados por suas parcerias futuras.

4. RESULTADOS ESPERADOS E IMPACTO

A ferramenta Bambu Enterprise tem o potencial de auxiliar significativamente democratizando o acesso a ferramentas avançadas na área do drug discovery, graças ao seu livre acesso a pequenas empresas, instituições acadêmicas e universitários que não possuem podem adquirir plataformas onerosas. A plataforma pode acelerar a descoberta de medicamentos para doenças que não possuem tanto pesquisa sendo consideradas negligenciadas, contribuindo para melhorar a saúde pública. O Bambu Enterprise é uma ferramenta desenvolvida como um dentro da universidade que será distribuído em parceria pela startup BioScient (https://www.bioscient.com.br/). A expansão para o mercado internacional, integração de novos modelos para outras doenças e a ampliação de sua base de usuários por meio de redes internacionais e parcerias são o foco atual para a implementação da ferramenta.

Os métodos empregados mostraram-se eficazes na otimização da descoberta de novos medicamentos para o tratamento de câncer. A integração das técnicas ETL com dados de NCBI e PubChem possibilitou uma organização robusta e padronizada das informações sobre compostos químicos. A automação do processo de ETL acelerou a análise e melhorou a qualidade dos dados, facilitando a triagem e validação de compostos. Essa abordagem não só otimizou o tempo e os recursos, mas também contribuiu para avanços na pesquisa e desenvolvimento de medicamentos, promovendo um progresso mais rápido e preciso na medicina personalizada.

5. CONCLUSÕES

A plataforma se demonstrou bastante inovadora devido facilitar o processo de drug discovery tornando mais acessíveis o uso dos modelos QSAR por meio de uma interface simples e colaborativa, retirando as questões financeiras e técnicas para pesquisadores. O foco do impacto social é a descoberta de novos medicamentos para beneficiar populações que sofrem com doenças consideradas negligenciadas. Devido à estratégia de distribuição e acesso, a ferramenta se

torna bem posicionada para cada vez mais se tornar popular e evoluir com novas funções para o seu uso.

Com todos esses problemas focados em serem resolvidos pela ferramenta, isso a torna um grande investimento para stakeholders, investidores e novos parceiros para se juntarem e acreditarem nesse projeto que demonstra ser promissor ao mercado de trabalho, como já possui uma propriedade intelectual em fase final para o seu registro, colaboração com a BIoScient, estamos abertos para conversas para estabelecermos relações tanto financeiras e acadêmicas para podermos expandir as oportunidades, objetivando avançar para as próximas etapas de desenvolvimento tanto do software quando no ambiente empresarial.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ATTENE-RAMOS, M. S.; AUSTIN, C. P.; XIA, M. High Throughput Screening. Em: WEXLER, P. (Ed.). Encyclopedia of Toxicology (Third Edition). Oxford: Academic Press, 2014. p. 916–917.

Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Front Pharmacol. 2018;9:1275. doi:10.3389/fphar.2018.01275.

Souibgui M, Atigui F, Zammali S, Cherfi S, Ben Yahia S. Data quality in ETL process: A preliminary study. Procedia Comput Sci. 2019;159:676-687. doi:10.1016/j.procs.2019.09.223.

Guidotti IL, Neis A, Martinez DP, Seixas FK, Machado K, Kremer FS. Bambu and its applications in the discovery of active molecules against melanoma. J Mol Graph Model. 2023;124:108564. doi:10.1016/j.jmgm.2023.108564.