Danh sách **100 chủ đề hấp dẫn nhất về ứng dụng dữ liệu ChEMBL\_35**, được trình bày song ngữ. Các chủ đề này mở rộng sang các lĩnh vực **AI, thiết kế phân tử, trực quan hóa và tích hợp dữ liệu đa nguồn**.

**🧪 I. Drug Discovery & Lead Optimization | Khám phá thuốc & tối ưu hóa lead (1–20)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 1 | QSAR model building from ChEMBL IC50 data | Xây dựng mô hình QSAR từ dữ liệu IC50 trong ChEMBL |
| 2 | Predicting bioactivity from molecular descriptors | Dự đoán hoạt tính sinh học từ các đặc trưng phân tử |
| 3 | Extracting lead-like compounds | Trích xuất các hợp chất giống lead |
| 4 | Filtering rule-of-five compliant molecules | Lọc các phân tử tuân thủ quy tắc Lipinski |
| 5 | Analyzing drug-likeness using RDKit | Phân tích tính giống thuốc bằng RDKit |
| 6 | Virtual screening based on ChEMBL targets | Sàng lọc ảo dựa trên các target trong ChEMBL |
| 7 | Ligand-based drug design using ChEMBL | Thiết kế thuốc dựa trên ligand từ ChEMBL |
| 8 | Structure-Activity Relationship (SAR) mining | Khai thác mối quan hệ cấu trúc – hoạt tính (SAR) |
| 9 | Tanimoto similarity search using ChEMBL + RDKit | Tìm kiếm tương đồng Tanimoto với ChEMBL và RDKit |
| 10 | Chemical space mapping of ChEMBL molecules | Bản đồ không gian hóa học của các phân tử ChEMBL |
| 11 | Predicting LogP and TPSA of ChEMBL compounds | Dự đoán LogP và TPSA của các hợp chất ChEMBL |
| 12 | Screening for CNS-active drugs | Sàng lọc thuốc tác động hệ thần kinh trung ương |
| 13 | Scaffold hopping analysis | Phân tích thay đổi khung hóa học (scaffold hopping) |
| 14 | Prioritizing compounds for synthesis | Ưu tiên hợp chất cho quá trình tổng hợp |
| 15 | Designing focused libraries from ChEMBL | Thiết kế thư viện hóa học tập trung từ ChEMBL |
| 16 | Finding natural product-like molecules | Tìm các phân tử giống sản phẩm tự nhiên |
| 17 | Mining kinase inhibitors from ChEMBL | Khai thác chất ức chế kinase từ ChEMBL |
| 18 | Target prediction for orphan molecules | Dự đoán mục tiêu cho các phân tử không rõ target |
| 19 | Linking ChEMBL with DrugBank for repurposing | Kết nối ChEMBL với DrugBank để tái định hướng thuốc |
| 20 | Comparing ChEMBL scaffolds with FDA drugs | So sánh khung hóa học ChEMBL với thuốc FDA |

**🧬 II. Target-Based Analysis | Phân tích theo đích tác dụng (21–40)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 21 | Top 50 protein targets by activity count | 50 protein mục tiêu có nhiều hoạt tính nhất |
| 22 | Analyzing GPCR-targeted ligands | Phân tích ligand nhắm GPCR |
| 23 | Extracting kinase-binding molecules | Trích xuất hợp chất gắn kinase |
| 24 | Target class distribution in ChEMBL | Phân bố các nhóm target trong ChEMBL |
| 25 | Multi-target ligand profiling | Lập hồ sơ ligand đa đích |
| 26 | Correlating bioactivity with target taxonomy | Liên hệ hoạt tính với phân loại target |
| 27 | Disease association of ChEMBL targets | Mối liên hệ giữa target và bệnh học |
| 28 | Ortholog mapping of targets (human → mouse) | Ánh xạ protein giữa người và chuột |
| 29 | Target pathway analysis via UniProt/KEGG | Phân tích pathway của target bằng UniProt/KEGG |
| 30 | Binding site comparison across targets | So sánh vị trí gắn ligand giữa các target |
| 31 | 3D target structure availability from PDB | Khả năng truy xuất cấu trúc 3D của target từ PDB |
| 32 | Predicting binding modes via molecular docking | Dự đoán cách thức gắn ligand bằng docking |
| 33 | Sequence similarity clustering of targets | Phân nhóm target theo tương đồng chuỗi |
| 34 | Heatmap of target-ligand interaction count | Biểu đồ nhiệt về số lượng ligand tương tác target |
| 35 | ChEMBL target network visualization | Trực quan hóa mạng lưới target |
| 36 | Target enrichment from gene sets | Làm giàu target từ bộ gene |
| 37 | Mining antimicrobial targets in ChEMBL | Khai thác target kháng khuẩn từ ChEMBL |
| 38 | Visualization of protein families | Trực quan hóa họ protein từ ChEMBL |
| 39 | Finding targets with dual antagonist/agonist activity | Tìm target có hoạt tính đối kháng và đồng vận |
| 40 | Top viral/bacterial targets in ChEMBL | Các target virus/vi khuẩn phổ biến nhất |

**💊 III. Clinical & Translational Insight | Thông tin lâm sàng & ứng dụng chuyển giao (41–60)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 41 | Linking ChEMBL drugs to clinical phases | Liên kết dữ liệu thuốc với các pha lâm sàng |
| 42 | Drugs with most activity data | Thuốc có nhiều dữ liệu hoạt tính nhất |
| 43 | Predicting clinical success from early bioactivity | Dự đoán khả năng thành công lâm sàng |
| 44 | Time-to-market estimation using ChEMBL history | Ước lượng thời gian ra thị trường từ ChEMBL |
| 45 | Mapping ChEMBL drugs to WHO ATC codes | Ánh xạ thuốc ChEMBL sang mã ATC của WHO |
| 46 | Drug repurposing candidates from ChEMBL | Các thuốc tiềm năng tái định hướng |
| 47 | Comparing clinical vs. preclinical molecules | So sánh thuốc lâm sàng và tiền lâm sàng |
| 48 | Linking ChEMBL with clinicaltrials.gov | Liên kết ChEMBL với dữ liệu thử nghiệm lâm sàng |
| 49 | FDA-approved ChEMBL drugs analysis | Phân tích thuốc ChEMBL đã được FDA phê duyệt |
| 50 | Trends in target class approval | Xu hướng phê duyệt theo nhóm target |
| 51 | Withdrawn drugs and their ChEMBL profiles | Hồ sơ thuốc bị rút khỏi thị trường |
| 52 | Oncology-focused drug trend in ChEMBL | Xu hướng thuốc ung thư trong ChEMBL |
| 53 | Antiviral compounds mapping to COVID-19 targets | Phân tử kháng virus tương ứng với đích COVID-19 |
| 54 | Neuropsychiatric drug insights from ChEMBL | Phân tích thuốc thần kinh – tâm thần |
| 55 | Drug safety flags and bioactivity profile | Cảnh báo an toàn liên kết với hoạt tính sinh học |
| 56 | Comparing ChEMBL drugs with WHO Essential Medicines | So sánh ChEMBL với danh mục thuốc thiết yếu WHO |
| 57 | Pharmacovigilance risk signals in ChEMBL | Dữ liệu tín hiệu cảnh báo dược lý |
| 58 | Pediatric drug insights from ChEMBL | Phân tích thuốc dành cho trẻ em |
| 59 | Orphan drug discovery via ChEMBL | Khám phá thuốc hiếm từ ChEMBL |
| 60 | ChEMBL → Real-world evidence mapping | Liên kết dữ liệu ChEMBL với dữ liệu thực tế |

**🤖 IV. AI & Machine Learning for Drug Discovery | Trí tuệ nhân tạo trong khám phá thuốc (61–80)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 61 | Building Random Forest QSAR models from ChEMBL | Xây dựng mô hình QSAR Random Forest từ ChEMBL |
| 62 | Deep learning for activity prediction | Học sâu để dự đoán hoạt tính sinh học |
| 63 | Feature selection for bioactivity classification | Chọn đặc trưng cho phân loại hoạt tính |
| 64 | Model interpretability with SHAP values | Giải thích mô hình bằng giá trị SHAP |
| 65 | Comparing ML algorithms on ChEMBL datasets | So sánh thuật toán ML trên dữ liệu ChEMBL |
| 66 | Predicting ADMET properties with ML | Dự đoán đặc tính ADMET bằng ML |
| 67 | Time-series trend analysis in bioactivity data | Phân tích xu hướng thời gian trong dữ liệu hoạt tính |
| 68 | Active learning strategies for compound selection | Chiến lược học chủ động trong chọn hợp chất |
| 69 | Model transferability across target classes | Khả năng chuyển mô hình giữa các nhóm đích |
| 70 | Ensemble models for robust bioactivity prediction | Mô hình tập hợp cho dự đoán hoạt tính ổn định |
| 71 | Cross-validation strategies for ChEMBL data | Chiến lược kiểm định chéo cho dữ liệu ChEMBL |
| 72 | Data augmentation with SMILES enumeration | Tăng cường dữ liệu bằng biến thể SMILES |
| 73 | Multitask learning on bioactivity matrix | Học đa nhiệm trên ma trận hoạt tính |
| 74 | AI-assisted target deconvolution | AI hỗ trợ phân rã đích tác dụng |
| 75 | Federated learning for privacy-preserving drug discovery | Học liên kết bảo vệ riêng tư trong khám phá thuốc |
| 76 | Using transformers for SMILES-based predictions | Dùng mô hình transformer để dự đoán từ SMILES |
| 77 | Fine-tuning ChemBERTa on ChEMBL data | Tinh chỉnh ChemBERTa trên dữ liệu ChEMBL |
| 78 | Molecule generation using generative models (VAE, GAN) | Tạo phân tử bằng mô hình sinh (VAE, GAN) |
| 79 | Using Gemini/GPT for drug data curation | Dùng Gemini/GPT để chuẩn hóa dữ liệu thuốc |
| 80 | Integrating ML models into Streamlit dashboards | Tích hợp mô hình học máy vào dashboard Streamlit |

**🧫 V. Molecular Informatics & Design | Tin sinh học phân tử và thiết kế hóa học (81–90)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 81 | Calculating molecular fingerprints (ECFP, MACCS) | Tính vân tay phân tử (ECFP, MACCS) |
| 82 | Generating 2D & 3D molecular structures | Tạo cấu trúc phân tử 2D & 3D |
| 83 | Fragment-based compound decomposition | Phân tích phân tử thành các mảnh hóa học |
| 84 | Enumerating tautomers and stereoisomers | Liệt kê các đồng phân tautomer và lập thể |
| 85 | Lipophilicity and solubility estimation | Ước tính độ tan và độ ưa lipid |
| 86 | Synthetic accessibility score (SAS) calculation | Tính điểm dễ tổng hợp của phân tử |
| 87 | PAINS filter implementation using RDKit | Lọc các phân tử nhiễu PAINS bằng RDKit |
| 88 | Molecule standardization and salt removal | Chuẩn hóa phân tử và loại muối |
| 89 | Scaffold extraction and Bemis–Murcko analysis | Trích xuất khung và phân tích Bemis–Murcko |
| 90 | Matched molecular pair analysis (MMPA) | Phân tích cặp phân tử tương đồng (MMPA) |

**📊 VI. Visualization & Web Integration | Trực quan hóa & tích hợp web (91–95)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 91 | Visualizing ChEMBL molecules in Streamlit | Trực quan hóa phân tử ChEMBL trong Streamlit |
| 92 | Plotting bioactivity heatmaps and clusters | Vẽ biểu đồ nhiệt và phân cụm hoạt tính |
| 93 | Creating interactive dashboards with Plotly | Tạo dashboard tương tác với Plotly |
| 94 | Embedding RDKit viewer in web app | Nhúng RDKit viewer trong ứng dụng web |
| 95 | Deploying ChEMBL explorer with FastAPI | Triển khai công cụ tra cứu ChEMBL bằng FastAPI |

**🌐 VII. External Dataset Integration | Tích hợp dữ liệu từ nguồn ngoài (96–100)**

| **STT** | **English Topic** | **Chủ đề tiếng Việt** |
| --- | --- | --- |
| 96 | Linking ChEMBL with PubChem CID/SMILES | Liên kết ChEMBL với CID/SMILES của PubChem |
| 97 | Enriching ChEMBL with DrugBank pharmacokinetics | Làm giàu ChEMBL với dữ liệu dược động học từ DrugBank |
| 98 | Mapping ChEMBL IDs to UniProt/ENSEMBL targets | Ánh xạ ID ChEMBL sang UniProt/ENSEMBL |
| 99 | Cross-referencing ChEMBL to PDB binding structures | Liên kết dữ liệu ChEMBL với cấu trúc gắn ligand từ PDB |
| 100 | Integrating ChEMBL with Open Targets platform | Tích hợp dữ liệu ChEMBL với nền tảng Open Targets |