

| | |
|-----------------|--|
| July 4, 2025 | <p>Project Initiation</p> <p>Began brainstorming topics for science fair project. Initially interested in neurodegenerative diseases, particularly Parkinson's disease, due to its increasing prevalence and lack of curative treatments.</p> |
| July 5, 2025 | <p>Background Research</p> <p>Researched mechanisms of Parkinson's disease including neuronal degeneration, dopamine loss, and neuroinflammation. Identified inflammation as a key area of interest.</p> |
| July 10, 2025 | <p>Exploration of Molecular Targets</p> <p>Investigated proteins involved in neurodegeneration. Discovered Matrix Metalloproteinase-9 (MMP-9) and its role in blood-brain barrier disruption and inflammation.</p> |
| July 14, 2025 | <p>Refining Research Focus</p> <p>Decided to focus on MMP-9 inhibition as a therapeutic strategy. Began exploring how drugs are traditionally discovered and the limitations of those methods.</p> |
| July 18, 2025 | <p>Introduction to Computational Drug Discovery</p> <p>Learned about in silico drug discovery, including QSAR models, virtual screening, and molecular docking. Recognized potential to apply machine learning.</p> |
| July 22, 2025 | <p>Project Direction Finalized</p> <p>Formulated idea to build a computational pipeline to identify potential MMP-9 inhibitors relevant to Parkinson's disease.</p> |
| July 27, 2025 | <p>Initial Method Planning</p> <p>Outlined phases of project: data collection, preprocessing, modeling, filtering, docking, and optimization.</p> |
| August 2, 2025 | <p>Learning Python for Scientific Computing</p> <p>Began working with Python libraries including pandas and NumPy for data handling.</p> |
| August 6, 2025 | <p>Introduction to RDKit</p> <p>Learned how to process chemical structures and calculate descriptors using RDKit.</p> |
| August 11, 2025 | <p>Dataset Exploration</p> <p>Investigated chemical databases. Selected ChEMBL as the primary source for MMP-9 inhibitor data.</p> |
| August 15, 2025 | <p>Dataset Acquisition</p> <p>Downloaded dataset containing 1067 compounds with SMILES and pIC50 values.</p> |
| August 20, 2025 | <p>Initial Data Inspection</p> |

| | |
|--------------------|---|
| | Reviewed dataset structure and identified missing values and formatting inconsistencies. |
| August 24, 2025 | Data Cleaning Removed compounds with missing SMILES or activity values. Ensured dataset completeness. |
| August 29, 2025 | Deduplication Checks Verified that no duplicate compounds existed using both IDs and SMILES strings. |
| September 3, 2025 | Descriptor Calculation Calculated molecular descriptors (MW, LogP, HBD, HBA) for all compounds. |
| September 7, 2025 | Fingerprint Generation Generated ECFP4 fingerprints to represent molecular substructures. |
| September 12, 2025 | Feature Matrix Construction Combined descriptors and fingerprints into a single dataset for modeling. |
| September 16, 2025 | Target Variable Definition Converted pIC50 values into binary classification (active vs inactive). |
| September 20, 2025 | Initial Model Training Built first Random Forest model and tested baseline performance. |
| September 24, 2025 | Model Evaluation Observed strong predictive performance (ROC-AUC ~0.92). Began analyzing results. |
| October 1, 2025 | Hyperparameter Optimization Used GridSearchCV to improve model performance and stability. |
| October 5, 2025 | Model Refinement Tested different parameter combinations and finalized optimized model. |
| October 10, 2025 | Model Validation (Y-Scrambling) Performed robustness test to confirm predictions were not random. |
| October 15, 2025 | Feature Importance Analysis Identified key features influencing predictions, including MW and LogP. |
| October 20, 2025 | Drug-Likeness Filtering Applied Lipinski's Rule of Five and Veber criteria to remove unsuitable compounds. |
| October 25, 2025 | Applicability Domain Analysis Ensured predictions were limited to relevant chemical space using similarity metrics. |

| | |
|-------------------|---|
| October 30, 2025 | Candidate Selection Filtered high-confidence compounds based on probability thresholds. |
| November 4, 2025 | Efficiency Metrics Evaluation Calculated ligand efficiency (LE) and lipophilic ligand efficiency (LLE). |
| November 9, 2025 | Multi-Objective Optimization Applied Pareto analysis to balance activity, stability, and drug-likeness. |
| November 14, 2025 | Final Candidate Shortlisting Selected top candidates for structural analysis. |
| November 20, 2025 | 3D Structure Generation Converted SMILES to 3D conformations. |
| November 25, 2025 | Protein Preparation Prepared MMP-9 structure for docking simulations. |
| December 1, 2025 | Molecular Docking Performed docking simulations to evaluate binding affinity. |
| December 5, 2025 | Selectivity Testing Tested compounds against related proteins (MMP-2, MMP-13). |
| December 10, 2025 | Interaction Analysis Analyzed binding interactions, including hydrogen bonds and zinc coordination. |
| December 15, 2025 | Lead Identification Identified Compound C3 as top candidate based on activity and pharmacokinetics. |
| December 20, 2025 | Pharmacokinetic Analysis Evaluated predicted half-life, clearance, and protein binding. |
| January 5, 2026 | Result Interpretation Analyzed trends across dataset and model outputs. |
| January 12, 2026 | Writing Results Section Began drafting detailed results and discussion. |
| January 20, 2026 | Analysis Development Connected results to methodology and evaluated strengths and limitations. |
| February 1, 2026 | Conclusion Drafting Summarized findings and implications for drug discovery. |
| February 10, 2026 | Final Report Editing Refined structure, clarity, and scientific tone. |

| | |
|-------------------|---|
| | |
| February 20, 2026 | Final Review Prepared project for submission. |
| March 1, 2026 | Presentation Worked on Presentation slides. |