# Project Readme Template

# Version 1 9/11/24

A single copy of this template should be filled out and submitted with each project submission, regardless of the number of students on the team. It should have the name readme\_"teamname" Also change the title of this template to "Project x Readme Team xxx"

Also change the title of this template to Project x Readine Team xxx			
1	Team Name: chealy5		
2	Team members names and netids: Catherine Healy (chealy5)		
3	Overall project attempted, with sub-projects: Implementing a polynomial time 2-SAT solver using DPLL algorithm		
4	Overall success of the project: Successfully coded program.		
5	Approximately total time (in hours) to complete: 10+ hours		
6	Link to github repository: <a href="https://github.com/CatherineH3/Theory">https://github.com/CatherineH3/Theory</a> Project 1 Final		
7	List of included files (if you have many files of a certain type, such as test files of different sizes, list just the folder): (Add more rows as necessary). Add more rows as necessary.		
	File/folder Name	File Contents and Use	
	Code Files		
	code 2sat chealy5.py	code 2sat chealy5.py is the code for my 2SAT solver. It can take an input file and generate a csv output with the times and satisfiability results. It also generates a graph.	
	Tes	st Files	
	check 1 chealy5.csv check 2SAT.cnf chealy5.csv check dumbsat (folder with check dumbsat chealy5.py and check dumbsat output chealy5.csv plots_DUMBSAT_chealy5.png)	check 1 chealy5.csv is a very short test with 3 simple cases that I created. I explain them in more detail in the testing section of the Readme. The only purpose of this is to verify correctness for a few cases by hand. It is not useful for timing because there are not enough test cases.  check 2SAT.cnf chealy5.csv was the provided test file. It contains 100 cases with clauses of varying lengths. The one	

file servers two purposes- it both is useful for verifying correctness and is useful also for timing. The main graph for the timing is actually generated in my main code, <a href="mailto:code\_2sat\_chealy5.py">code\_2sat\_chealy5.py</a>.

check dumbsat is a folder containing test files that I used to verify my solution. I generated a solution using dumbsat using the same input file as 2SAT and compared it to the solution using 2SAT to verify correctness and also for a time comparison.

<u>check dumbsat chealy5.py</u> is the code written to test the same input for DumbSAT

<u>check dumbsat output chealy5.csv</u> is the output to the <u>check dumbsat chealy5.py</u> code using input

<u>check 2SAT.cnf chealy5.csv.</u> The result is <u>plots DUMBSAT chealy5.png</u>

#### Data File

#### data cnf chealy5.csv

data\_cnf\_chealy5.csv\_ is\_actually identical to check\_2SAT.cnf\_chealy5.csv\_because I used it for both a timing run and for verifying correctness. I included it again under another name to follow the submission guidelines.

#### **Output Files**

output metadata chealy5.docx output chealy5.csv output test1 chealy5.csv Note: I could not add all of the necessary information into the csv file, so I added an additional metadata file for the output: output metadata chealy5.docx.

output\_chealy5.csv is the output after running the main 2Sat solver, code\_2sat\_chealy5.py, with data input: data\_cnf\_chealy5.csv. The data input is the same as the one provided as a test file. It is a csv with the runtime, number of literals, and whether it was satisfiable. This output was used for two purposes: (1) to verify the correctness of the code and (2) To record the timing of the code and show results.

	C ";	a screenshot of my screen after running the ode in Colabs is also included in the additional information" section of this ocument to prove that it actually ran.  autput_test1_chealy5.csv is the output after sg input check 1_chealy5.csv.	
	Plots (as needed)		
	plots run1_chealy5.png plots run2_chealy5.png plots run3_chealy5.png	Plots_run 1-3 are all plots as a result of running code_2sat_chealy5.py with data_cnf_chealy5.csv input file. Each plot graphs the time vs. the number of literals. I included 3 cases because the results vary.	
	Other files		
	output console screenshot chealy5.png	output_console_screenshot_chealy5.png is a screenshot to prove I actually ran the code	
8	Programming languages used, and associated libraries: For this project, I used the python programming language. I used the following libraries: time, csv, dataclasses, matplotlib.pyplot, and numpy. I used the time library to calculate how long it took to solve each wff SAT problem. I used the dataclasses library to create my own dataclass for each wff to store its information. I used matplotlib.pyplot to plot my results. I used numpy to calculate the trend line.		
9	Key data structures (for each sub-project): The key data structure that I used was a dataclass to hold the information of each wff. It is as described below. class WFF:     problem_number: int # Problem Number     n_variables: int # Number of variables     satisfiability: str # S for satisfieable, U for unsatisfiable, ? for unknown     all_clauses: list[list[int]] # store the clauses as a list of ints     n_clauses: int = 0 # Number of clauses (initialized ot 0)     time_to_solve: int = 0 # Keep track of how long it took to solve     n_literals: int = 0 # number of literals is number of clauses * 2 Having this dataclass was very helpful because it provided a clear way to store all the information about each wff in an organized manner (ex: mywffs was a list I used), and it made it very easy to generate the output files. I also used many lists to keep track of all my lists. Using lists was helpful to stay organized and print all of my information. Additionally, I used a dictionary in the remove_pure_literals function in my DPLL algorithm to quickly count the number of each literal and try to identify whether or not there was a pure literal (a pure literal exists without its negation, so I had to check them all).  Lastly, although I did not directly interact with the stack and use stack operations, I		

made use of the stack through my DPLL algorithm. Because the algorithm was recursive, it internally used the stack to keep track of the changing variable assignments when guessing each possible variable assignment.

10 | General operation of code (for each subproject)

My project was the polynomial time 2SAT solver. The code takes in an input csv file in cnf form and generates an output csv file that containts the number of literals, execution time, and runtime fore each wff. It does not accept files in other formats. In the main, I focused on parsing the csv file and putting all of the information into a WFF dataclass. Once it is all parsed and all the WFF dataclasses are stored in a list, the satSolve function is called. This function loops through each wff, calls the dpll algorithm to solve it, for each one. After seeing the result of whether it is satisfiable or unsatisfiable, it stores it in the WFF dataclass. Additionally, the function measures the amount of time it takes for the DPLL algorithm to run using the time library and saves this information. The DPLL function calls two other functions- one for unit propagation and another for elimination of pure literals. First, it searches for any unit clauses and calls the unit propagation function. This function simplifies the clause based on the literal assignment. Then, the DPLL function calls a function to remove pure literals. This searches for any literals that occur without their negation and removes them. Then, it checks if there are no more literals and if so, return true. If the length of the clause is 0, return false because it is unsatisfiable because there are no literals left to choose. The algorithm then chooses the next literal. If there are no more literals left to choose, it means it is satisfiable and returns true. Lastly, but importantly, there is the recursive call for two cases. One case assumes the literal is true and propagates. The other assumes the literal is false and propagates.

After calling DPLL, the main also calls a function to print the output. The output file is in the form of a csv. It was easy to print because I stored all of the information in a dataclass.

After this, a plotting function is called. The plotting function uses Matplotlib. I also generated a trendline that was a polynomial trendline (because this is a polynomial time solver). I initially assigned it to degree two and used numpy functions to do this.

What test cases you used/added, why you used them, what did they tell you about the correctness of your code.

For an initial test, I made my own cnf file with 3 cases:

c,1000,2,?

p,cnf,2,3

-2,4,0,

-2,4,0, -2,4,0

2,-4,0

c,1001,2,?

p,cnf,2,4

-2, -3, 0,

3,2,0,

-2,3,0,

2,-3,0,

c,1002,3,?

p,cnf,3,4

4,5,0,

6,-5,0,

-4,5,0, 5,-6,0,

The first one is definitely satisfiable because the assignment 2 = True and 4 = True leads to a True result overall.

The second case is unsatisfiable because it is composed of every combination of the two variables, so it would be impossible there to be an assignment that leads to a True result overall. If they both are true, it is false. If the both are false, it is false. If the 3 is true and the 2 is false, it is false. If the 2 is true and the 3 is false it is false.

The final case is satisfiable because the assignment 4 = T 5 = T 6 = T would make the overall wff true.

I ran the code on these and it got the correct results.

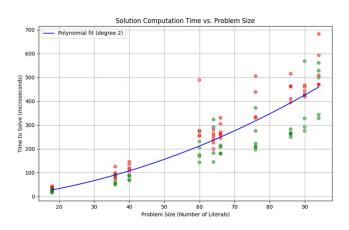
For more robust testing, I used the provided test cases that were given in Canvas for my primary testing. The canvas file provided: 2SAT.cnf.csv, was renamed to fit the name convention. This has 100 test cases of varying sizes, so I was considered this to be sufficient testing. To determine the correct answer, I actually ran the same exact input through DumbSAT and compared the results. Because I was told that DumbSAT was a correct algorithm and it was provided, this was very helpful to use as a comparison. However, the downside is that DumbSAT took about 1 hour and 15 minutes to run on my computer. DumbSAT was initially set up to generate its own test cases randomly, so I had to make some modifications. I took the code I used from my 2SAT solver to parse the input file and called the DumbSAT function. With my WFF dataclass, it was easy to give Dumbsat all of the information it needed to run. Then, I used thhe same function I wrote from the 2SAT solver that outpts the information to output DumbSAT results. This output file has the number of literals, the computation time, and the satisfiability. After directly comparing satisfiability with DumbSAT, I found them to be identical for all 100 test cases. I felt that 100 test cases was sufficient to consider my output correct. Additionally, my graph appeared to follow the expected trends (discussed in the result discussion section), so I felt pretty good about the correctness.

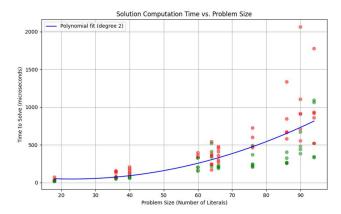
### 12 How you managed the code development

I completed all of the code myself because I was an individual team. To develop the code, I first focused on developing code to read in the file. I originally ran my code in the terminal using VScode. I later switched to Google Collabs because I was unsure if I could graph in Vscode, and I knew that I had graphed in google collabs before. A major first step in the development of the code was developing the code to read in data from the csv file that was in cnf format. I spent a long time trying to understand what this format was saying. In order the read it in, I ended up making a dataclass for the wff, which allowed me to store lots of information easily about the wff including the number of literals, the number of variables, the runtime, the clauses, and the satisfiability itself. This was very useful because it made it very easy to generate the output files because I stored the information in an organized manner. Then, I developed code to print the results and generate the result file. Then, to verify that it worked, I actually tested it by running my code with DumbSAT algorithm that was provided. I modified it slightly so it would work with the input in the cnf form instead of generating its own test cases like it did in the provided file. After verifying that it worked, I worked on the DPLL algorithm. I was not familiar with this algorithm at all, so I looked online for more information. The wikipedia page was very helpful: https://en.wikipedia.org/wiki/DPLL algorithm After looking at various pseudocode, I attempted to implement it. To debug, I wrote many print statements printing out the information that I had so I could trace where the

error was occurring. I timed the code within the functions and stored all the info, so it was very easy to generate the output files. I then used matplotlib to generate the graph. For testing, I thought of a few test cases and ran them. My main was of testing was to actually run the same input (the provided test cases) through DumbSAT, which I had modified a little, and this created a solution that I could compare my code with.

#### 13 Detailed discussion of results:





My results were accurate because I checked by running the exact same code through DumbSAT for a comparison. In general, the DPLL algorithm in the worst case is O(2<sup>n</sup>), where n is the number of variable in the formula. However, for 2SAT, a special case, it is expected to be more linear for most cases because the clauses are short and this leads to a lot of logical shortcuts through the unit propagation and pure literal elimination techniques.

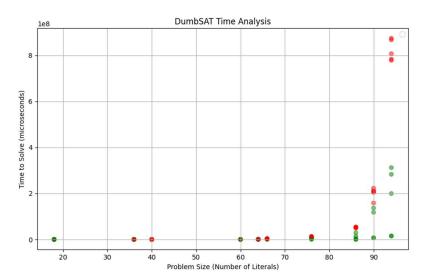
The 2-Sat problem can be solved in polynomial time. It can be solved in O(n+m), where n is the number of variables and m is the number of clauses. It is polynomial time, even in the worst case. It is possible to get linear. This approach is P (Not NP) because it is in polynomial time. 2-SAT is faster because if one literal is false, the other must be true. There is a domino effect and it is easier to determine assignments than other cases.

I generated multiple plots because the exact computation time varies a little, so the curvature varies. However, it is clear that you can see that as the number of literals

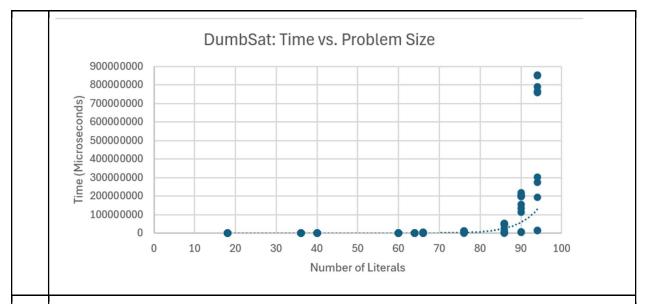
increase, the computation time directly increases. One of my plots does appear to be pretty close to linear, while the other is slightly more curved. I used a polynomial degree trendline to be fit to my graph and it appears to be a good fit, so this confirms that my solution was in polynomial time.

As a comparison, I also graphed DumbSAT. DumbSAT is actually exponential and is O(2<sup>variables</sup> \* clauses \* average literals). As the number of literals gets bigger, it slows down completely. DumbSAT took over an hour to run on my computer, while my 2SAT solver took a couple of seconds. Clearly, my algorithm is a good speedup.

Unfortunately, I didn't get to add a trend line to Dumbsat because I was having trouble figuring out how to make an exponential fit and it took several hours each time I reran the code. However, you can clearly see that it is exponential and that it goes up very steeply as the problem size gets bigger. Towards the end, the program took many minutes to run and the program as a whole took hours even though the input was identical to the 2SAT input.



In order to view the data with an exponential trendline, I quickly graphed it in Excel but couldn't figure out how to color it based on satisfiable / unsatisfiable. The exponential trendline seemed to fit pretty well.



# 14 How team was organized

Because I was an individual team (a team of 1), I completed all of the work myself. I was responsible for doing the main code to read in the input, I created the DPLL algorithm code, and I generated the output. I also generated the graphs and the tests. I also modified Dumbsat for a comparison. I was in charge of verifying everything. Overall, the organization was mostly good. One area of improvement is that I switched my project 3 different times, so I would try to organize better and be more decisive in the future.

What you might do differently if you did the project again

If I did this project again, I would definitely start the project earlier. It was very stressful to wait until the last minute to start it, and it was difficult to rush the process. Additionally, I would take more time when selecting which project I would do. I switched projects 3 times out of indecisiveness and wasted a lot of time. Lastly, I would try harder to understand the DPLL algorithm first before coding it. I jumped into coding it too quickly without fully understanding, and it took a while to figure it out and debug.

## 16 Any additional material:

Professor Kogge mentioned including a screenshot of our console to prove that the code actually ran. I am not sure where to include this, so I will include it here:

