Task 1

To prove that the relations R_1 and R_2' are isomorphic, we need to demonstrate that there exists a bijective function $f : A \to B$ such that $(a, b) \in R_1$ if and only if $f(a), f(b) \in R_2'$.

Given

- Set $A = \{1, 2, 3, 4, 5, 6\}$ with a relation R_1 such that $(a, b) \in R_1$ if and only if $a + b$ is even.
- Set $B = \{'a', 'b', 'c', 'd', 'e', 'f'\}$ with a relation R_2' such that $(x, y) \in R_2'$ if and only if the alphabetical distance between x and y is even.

Solution - Isomorphism

Define a bijective function $f : A \to B$. For simplicity, let's choose f as:

$$f(1) =' a'$$
$$f(2) =' b'$$
$$f(3) =' c'$$
$$f(4) =' d'$$
$$f(5) =' e'$$
$$f(6) =' f'$$

Now, let's verify the relations:
For R_1: $(a, b) \in R_1$ if and only if $a + b$ is even.
For R_2': $(x, y) \in R_2'$ if and only if the alphabetical distance between x and y is even.
For $(a, b) \in R_1$, their sum is even if:

- both are odd: Possible pairs are (1, 3), (1, 5), (3, 5), (3, 1), (5, 1), (5, 3)
- both are even: Possible pairs are (2, 4), (2, 6), (4, 6), (4, 2), (6, 2), (6, 4)

Now, consider the pairs in B:
Alphabetical distance is even when both characters are even or both are odd. This implies:

- Both are odd: ('a', 'c'), ('a', 'e'), ('c', 'e'), ('c', 'a'), ('e', 'a'), ('e', 'c')
- Both are even: ('b', 'd'), ('b', 'f'), ('d', 'f'), ('d', 'b'), ('f', 'b'), ('f', 'd')

Notice that both the sets of pairs match when we use our defined function f.
Thus, our function f preserves the relation.

Solution - Topology

For determining the topological sorting, we need to consider the relation R_1 on set A and R_2' on set B.

- For R_1:

  $$(a, b) \in R_1 \text{ if and only if } a + b \text{ is even.}$$

  Therefore, for any two even or two odd numbers, their sum will be even, and thus, there will be an edge between them in the graph. This graph will be undirected, as the relation is symmetric. Therefore, a topological sort is not possible for the graph of R_1.

- For R_2':

  $$(x, y) \in R_2' \text{ if and only if the alphabetical distance between } x \text{ and } y \text{ is even.}$$

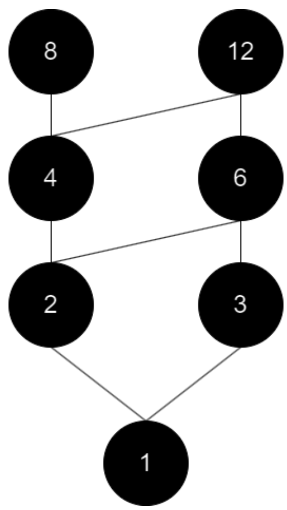
  Similar to R_1, the graph for R_2' will also be undirected, as the relation is also symmetric. Therefore, a topological sort is also not possible for the graph of R_2'.

Conclusion: A topological sort is not possible for the graphs of both relations.

Task 2

The set is $P = \{1, 2, 3, 4, 6, 8, 12\}$ and the relation $\leq$ is based on divisibility.

$$\begin{aligned} &1 : 2, 3, 4, 6, 8, 12 \\ &2 : 4, 6, 8, 12 \\ &3 : 6, 12 \\ &4 : 8, 12 \\ &6 : 12 \end{aligned}$$



Antichains

An antichain is a subset of a poset in which no two elements are comparable.

- {2, 3}
- {3, 4}
- {2, 4, 6}

Order Ideals

An order ideal is a subset I of a poset such that if x is in I and y is less than x, then y is also in I.

- {1}
- {1, 2, 3}
- {1, 2, 4}

Order Filters

An order filter is a subset F of a poset such that if x is in F and y is greater than x, then y is also in F.

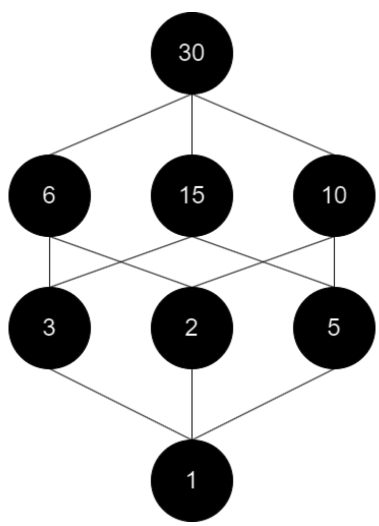
- {12}
- {6, 12}
- {4, 8, 12}

Task 3

Hasse Diagram for Poset (P, $\leq$)

The set is $P = \{1, 2, 3, 5, 6, 10, 15, 30\}$ and the relation $\leq$ is based on divisibility.

$$\begin{aligned}
1 &: 2, 3, 5, 6, 10, 15, 30 \\
2 &: 6, 10, 30 \\
3 &: 6, 15, 30 \\
5 &: 10, 15, 30 \\
6 &: 30 \\
10 &: 30 \\
15 &: 30
\end{aligned}$$



Order Dimension of the Poset (P, $\leq$)

The order dimension of a poset is the smallest number of total orders whose intersection is the given partial order. In simpler terms, it's about finding how many linear extensions (or linear orderings) we need such that, when combined, they reflect all the relations of our poset.

For the given poset $P = \{1, 2, 3, 5, 6, 10, 15, 30\}$, we can have the following linear extensions:

1. $1, 2, 3, 5, 6, 10, 15, 30$
2. $1, 3, 5, 2, 6, 15, 10, 30$

When combined, these linear extensions reflect all the relations of our poset. Thus, the order dimension of the poset $(P, \leq)$ is 2.

Why is this so?

Attempting to represent this partially ordered set with a single linear extension, we encounter the issue that we cannot account for all the "divides" relations. For instance, if we place 2 before 6, we would also need to place 2 before 3 to account for 6 (since 2 divides 6). However, 2 doesn't divide 3. Therefore, one linear extension is insufficient.

Now, let's consider two linear extensions:

In the first extension, we arrange the numbers to account for the divisibility relations for 2:

$$1, 2, 3, 5, 6, 10, 15, 30$$

In the second extension, we arrange the numbers to account for the divisibility relations for other numbers, like 3:

$$1, 3, 5, 2, 6, 15, 10, 30$$

Using these two linear extensions, we can cover all divisibility relations from the original partial order. Thus, the minimum number of linear extensions that collectively reflect all relations of the partial order is two.

Task 4

Dataset 1

- Dataset `https://archive.ics.uci.edu/dataset/53/iris`
- Target value - Classlabels (Classifying the type of iris)
- Head of frame -

Sepal length	Sepal width	Petal length	Petal width	Class_labels
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3.0	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5.0	3.6	1.4	0.2	Iris-setosa

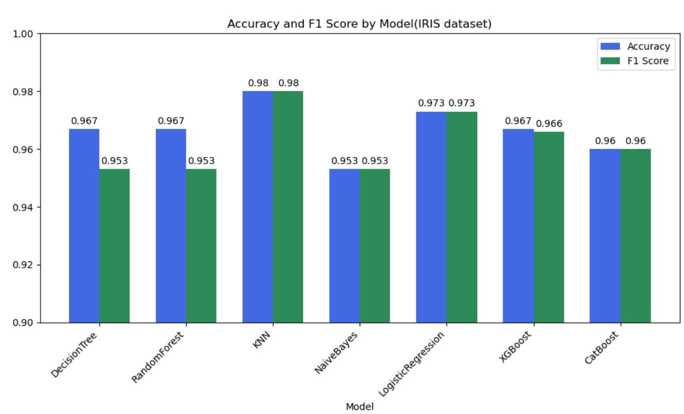
Table 1: DataFrame

- Model evaluation results -

Model	Accuracy	F1 Score
DecisionTree	0.967	0.953
RandomForest	0.967	0.953
KNN	0.980	0.980
NaiveBayes	0.953	0.953
LogisticRegression	0.973	0.973
XGBoost	0.967	0.966
CatBoost	0.960	0.960

Table 2: Model evaluation results

- Accuracy and F1 Score by Model(IRIS dataset) -



Dataset 2

- Dataset `https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset`
- Target value - Stroke (Stroke prediction)
- Head of frame -

Part 1

id	gender	age	hypertension	heart_disease	ever_married	work_type
9046	Male	67.0	0	1	Yes	Private
51676	Female	61.0	0	0	Yes	Self-employed
31112	Male	80.0	0	1	Yes	Private
60182	Female	49.0	0	0	Yes	Private
1665	Female	79.0	1	0	Yes	Self-employed

Part 2

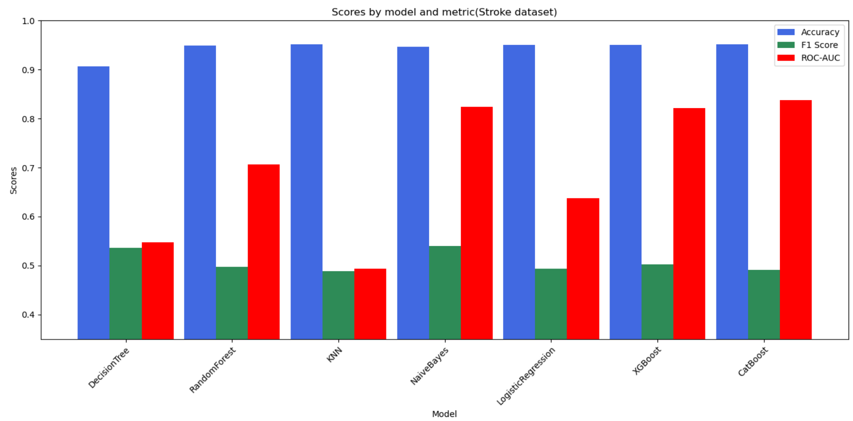
Residence_type	avg_glucose_level	bmi	smoking_status	stroke
Urban	228.69	36.6	formerly smoked	1
Rural	202.21	NaN	never smoked	1
Rural	105.92	32.5	never smoked	1
Urban	171.23	34.4	smokes	1
Rural	174.12	24.0	never smoked	1

- Model evaluation results -

Model	Accuracy	F1 Score	ROC-AUC
DecisionTree	0.906	0.536	0.547
RandomForest	0.949	0.498	0.706
KNN	0.951	0.488	0.494
NaiveBayes	0.946	0.540	0.824
LogisticRegression	0.950	0.494	0.638
XGBoost	0.950	0.502	0.821
CatBoost	0.951	0.491	0.838

Table 5: Model evaluation results

- Scores by model and metric(Stroke dataset) -



It seems we've encountered a typical problem when working with imbalanced datasets. If one class significantly outnumbers the other in instances (for example, the majority of patients don't have a stroke), many models might simply "ignore" the minority class, leading to high accuracy but a low F1 score.
There are many ways to deal with imbalanced datasets, but my attempts haven't been successful so far. Therefore, we'll add another metric, ROC-AUC.
ROC-AUC is a metric that can be more informative for imbalanced datasets than accuracy or the F1 score.

Dataset 3

- Dataset `https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction`
- Target value - Satisfaction (Satisfaction prediction)
- Head of frame - table too big
- Model evaluation results -

Model	Accuracy	F1 Score
DecisionTree	0.944	0.944
RandomForest	0.961	0.961
KNN	0.928	0.926
NaiveBayes	0.864	0.861
LogisticRegression	0.876	0.873
XGBoost	0.961	0.960
CatBoost	0.961	0.960

Table 6: Model evaluation results

- Scores by model and metric(Stroke dataset) -