**Slide 1:** Hi everyone! I’m Boti, and today I’m excited to share my paper pitch with you. The paper I chose is titled *Feature Selection for Support Vector Machines via Mixed Integer Linear Programming*, a study by Maldonado, Pérez, Weber, and Labbé from 2014.

**Slide 2:** I would like to start with a look at the structure of my presentation.

**Slide 3:** My plan is to briefly cover some key concepts from the paper’s title. First, the *support vector machine* (SVM) classifier method—probably familiar to most of you. Then, I’ll touch on *feature selection*, a fundamental yet critical concept in machine learning. Lastly, I’ll introduce *MILP*, or mixed integer linear programming, which is perhaps the most mathematical of the three and may be less familiar. After covering these basics, my presentation will follow this structure: **Motivation** – The importance of this research and the problem it aims to contribute to. **Existing Methods** – A look at previously established methods in SVM feature selection, which is extremely important for the following part. **Proposed Methods** – The new MILP-based approaches introduced in this paper. Last, but not least **Results** – How these new methods perform in practice, and what insights were gained.

**Slide 4:** Now, let’s move on to the motivation behind this paper.

**Slide 5:** While this study may seem niche, it’s actually quite valuable. This is because SVMs are among the most popular ML models, especially amongst classifiers. Feature selection, in turn, is crucial, but choosing the optimal subset of features is an *NP-hard problem*—in other words, solving it is computationally intense. This paper’s goal is to propose competitive new methods for feature selection in SVMs using MILP to tackle this complexity.

**Slide 6:** Next, I’ll give you an overview of the methods discussed in the paper.

**Slide 7**: For that just super quickly let me remind you that a support vector machine is a binary classifier, which aims to have the largest margin possible – that is the distance to the closest datapoint. For that we must solve a constrained optimization problem, which can be seen here, but let’s not get lost in the details.

**Slide 8**: Another important thing to mention is the concept of an embedded model. There are different ways to select the optimal features. To reduce computation time, it makes sense to come up with ways to train a model only once and select the optimal features during training. Such models are called embedded models.

**Slide 9:** Finally, let’s have a look at the list of the models related to SVM. On the left are existing methods referenced and used in the paper, while on the right are the paper’s new approaches, *MILP1* and *MILP2*. Let me quickly just say that from the existing methods the l\_1 method changed the optimizer function to have a Lasso penalty, other than that it uses the original formulation, while the LP method reformulated it in a way that linear programming can be used. I explained these 2, because the new methods adapt them to work with mixed integer programming.

**Slide 10:** Now a slide on what the paper achieved.

**Slide 11:** The new methods along with the existing ones were evaluated on 7 different datasets. On the picture above you can see results for the first 3, but if you can believe me, let me say that in all 7 cases one of the 2 methods achieved the best accuracy while having a reasonable number of features selected. However, one issue occurred, namely that while their runtimes also outperformed the other methods for the first datasets, but MILP2 starts to have massive computational times if the dataset set gets complicated.

**Slide 12:** I hope this gave you a clear, concise overview of my topic—thanks for your attention!