**DSML Project Ideas:**

1. Recommender System Projects

Have you ever seen movies or web series on online streaming platforms? Once you watch one or two of them, you will notice that apps like Netflix and Amazon Prime recommend new web series and movies. It is because these apps render Machine learning models that try to understand the customer's taste. Modern e-commerce sites like Flipkart, Amazon, Alibaba, etc., also have the same feature. Recommendation engines are popular in media, entertainment, and shopping. All modern apps come with a recommendation engine that suggests users for more engagement.

2. Building Chatbots

* Language: Python
* Data set: Intents JSON file
* Source code: [Build Your First Python Chatbot Project](https://dzone.com/articles/python-chatbot-project-build-your-first-python-pro)

Chatbots play a pivotal role for businesses as they can effortlessly without any slowdown. They [automate](https://builtin.com/articles/robotic-process-automation) a majority of the customer service process, single-handedly reducing the customer service workload. The [chatbots](https://builtin.com/design-ux/chatbot-turing-test-mitsuku-pandorabots) utilize a variety of techniques backed with artificial intelligence, [machine learning](https://builtin.com/machine-learning) and data science.

Chatbots analyse the input from the customer and reply with an appropriate mapped response. To train the chatbot, you can use [recurrent neural networks](https://builtin.com/data-science/recurrent-neural-networks-and-lstm) with the [intents JSON dataset](https://github.com/katanaml/katana-assistant/blob/master/mlbackend/intents.json), while the implementation can be handled using [Python](https://builtin.com/learn/tech-dictionary/python). Whether you want your chatbot to be domain-specific or open-domain depends on its purpose. As these chatbots process more interactions, their intelligence and accuracy also increase.

*(Intelligent Chatbots*

*Chatbots are intelligent systems or a program that can communicate and assist users similar to that of humans. All modern companies develop chatbots and integrate them into their web apps or smartphone apps to solve user queries and FAQs.*[*Developing chatbots*](https://www.projectpro.io/article/python-chatbot-project-learn-to-build-a-chatbot-from-scratch/429)*have become the need of the hour with almost 9 out of 10 companies using them for enhanced customer experience. So, creating an ML-based chatbot from scratch will be an excellent project for the final year students. There are two basic kinds of chatbot models:*

*·        Retrieval Based models and*

*·        Generative Based models*

*This project will require concepts like Natural Language Processing and artificial neural networks. In this project, you will need libraries like JSON, Natural Language Toolkit (*[*NLTK*](https://www.projectpro.io/article/nltk/846)*), Pickle, and frameworks like TensorFlow and Keras. Chatbots require different types of datasets like question-answer datasets, conversation datasets, logical reply-based datasets, etc. Some well-known datasets you can use to train your chatbot are:*

* [*Yahoo Language data*](https://webscope.sandbox.yahoo.com/catalog.php?datatype=i&did=67&guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuYW5hbHl0aWNzaW5zaWdodC5uZXQvdG9wLTEwLWNoYXRib3QtZGF0YXNldHMtYXNzaXN0aW5nLWluLW1sLWFuZC1ubHAtcHJvamVjdHMv&guce_referrer_sig=AQAAAG7sPSWqwFwtGKPEYEby6ESBX5CFuXuMJpXYKwCLApDyNbRumbHE1N8c7E5Dqxh4oaz2h41DEqd9tdEhDJTPgmfxr7rw-u5beiHU7fxBaJYxJ4SYv7kQpsciaW-Z5zhkpxiIhK3JPGOUuJd_5h2wU1w-F3lrtOczhfHo2bwgQm2X)
* [*Stanford Question Answering Dataset (SQuAD)*](https://rajpurkar.github.io/SQuAD-explorer/)
* [*ClariQ*](https://github.com/aliannejadi/ClariQ)
* [*NPS Chat Corpus*](http://faculty.nps.edu/cmartell/NPSChat.htm)
* [*HotpotQA*](https://hotpotqa.github.io/)
* [*Shaping Answers with Rules through Conversation (ShARC)*](https://arxiv.org/abs/1809.01494)

*Among them, NPS Chat Corpus comes as a component of Natural Language Toolkit (NLTK) distribution. Stanford Question Answering Dataset (SQuAD) is another comprehension-based dataset. It has a collection of well-researched questions professed by crowd workers that can help build an effective training model for your chatbot project.)*

3.Email Spam-Filtering System

Mining text is one of the popular computation techniques widely applied in applications like [text summarization](https://www.projectpro.io/article/text-summarization-projects/693), topic classification, machine translation, [sentiment analysis](https://www.projectpro.io/article/sentiment-analysis-project-ideas-with-source-code/518), etc. Modern [cybersecurity](https://www.projectpro.io/article/cybersecurity-machine-learning-projects/631) systems are utilizing machine learning methods a lot. Spam email detecting systems are one of them. Spam filtering also leverages [text mining](https://www.projectpro.io/article/text-mining-projects/755) and document classification to segregate legitimate mails and spam emails. All modernized email services come with this segregation system that runs machine learning algorithms behind. Such a project comes under the text [classification problems](https://www.projectpro.io/article/7-types-of-classification-algorithms-in-machine-learning/435). Building this kind of a ML project involves the following important steps -

a.      Text Processing

b.      Text Sequencing

c.      [Model Selection](https://www.projectpro.io/article/model-selection-in-machine-learning/824)

d.      Implementation

Use libraries like Sklearn, NumPy, Counter, Scrubadub, Beautifier, Seaborn, and machine learning frameworks like TensorFlow and Keras. For training such machine learning models, a[Spambase dataset](http://archive.ics.uci.edu/ml/datasets/Spambase/) can help you. Spambase dataset is an open-source UCI machine learning repository comprising around 5569 emails, of which nearly 745 are spam emails.

4.Credit Card Fraud Detection Project

Finding anomalies in credit cards and detecting fake credit card transactions was challenging before the advent of machine learning. This project will approve and differentiate swindling credit card transactions from legitimate ones. This project will make you discover and learn how to perform the classification of data. Also, you have to be thorough with the concepts like decision trees, artificial neural networks (ANN), logistic regression, and gradient-boosting classifiers. You can implement the [credit card fraud detection project](https://www.projectpro.io/project-use-case/anomaly-detection-using-deep-learning-and-autoencoders?utm_source=MLStuBlg&utm_medium=Prolink) using libraries like NumPy, Pandas, Matplotlib, Seaborn, XGBClassifier, and frameworks like Scikit-Learn. The[credit card dataset](https://datahub.io/machine-learning/creditcard#readme) and[credit-card fraud detection dataset](https://www.kaggle.com/mlg-ulb/creditcardfraud) are preferable to train your ML model for such a project. These datasets have credit card details along with dummy data of fraudulent and non-fraudulent transactions.

5. Fake News Detection Project

It is another innovative machine learning project for final year students. As you all know, fake news is speeding like a bushfire. Right from connecting people to reading the daily news, everything is available on social media. Hence, it has become more skeptical these days to detect fake news. Many popular social media platforms like Facebook and Twitter already have [fake news detection](https://www.projectpro.io/article/fake-news-detection-project/854) algorithms running behind the scenes on posts and feeds. Implementing this kind of ML project requires good know-how of [various NLP techniques](https://www.projectpro.io/article/10-nlp-techniques-every-data-scientist-should-know/415) and classification techniques (PassiveAggressiveClassifier or Naive Bayes classifier) to detect fake news. PassiveAggressiveClassifier is an online learning algorithm that remains passive while discovering correct classification outcomes. You can prefer supervised learning models to develop such projects.

Libraries like NumPy, Pandas, Itertools, and [spaCy](https://www.projectpro.io/article/spacy-projects-examples/624" \o "spaCy" \t "_blank) (for NLP tasks) will become handy for such a project. Apart from that, frameworks like Scikit-Learn and [Streamlit](https://www.projectpro.io/article/streamlit-python-projects/622" \o "Streamlit" \t "_blank) will help a lot. Scikit-Learn has different machine learning approaches and statistical [modeling](https://www.projectpro.io/article/data-modeling-projects/676data%20models" \o "modeling" \t "_blank) for clustering, classification, and regression. Streamlit helps in building web applications more efficiently and quickly and includes the deployment of ML models. Also, the[Great Fake News dataset](https://github.com/derevirn/gfn-detector/tree/main/data) and[ISOT Fake News Dataset](https://www.uvic.ca/engineering/ece/isot/datasets/fake-news/index.php) are the best for training this ML model.