

Chatbot Report

System Description:

This program is an implementation of a very simple and naive chatbot. This chatbot has a domain knowledge around the Los Angeles Lakers and can only speak regarding that subject. At times, the chatbot answers perfectly at other times not so much. This program uses a lot of techniques discussed in class from tokenization, POS tagging, term count, tfidf, and machine learning. The chatbot works like this. It introduces itself and gets the information of the user. The user information is stored and then allows the user the freedom to type anything. The program uses rule based and ML to determine if it is a question, if so, it tries to answer it by doing cosine similarity between the query and the knowledge base. It returns the most common sentence. If the input is not a question, then it determines if it is a command using rule based, if so it tries to do that command by treating it as a question. Otherwise, it performs sentiment analysis using ML to determine if it is an opinion that is negative or positive of the user. It then stores it with the connotation and replies to it. The process then repeats until the user enters a command that makes it exit.

Logical Steps (Simplified):

1. Create ML Model for Sentiment Analysis (NLP Tech: ML)
2. Introduce chatbot and store user information
3. Get user input and determine if it is a question (NLP Tech: POS Tagging & ML)
4. If question, then formulate an answer using the knowledge base (NLP Tech: Tokenization, TFIDF, Cosine Similarity)
5. If not a question, determine if it is a command or statement
 - a. If command, then treat like a question
 - b. If statement, predict sentiment using ML model in step 1 and reply based on the sentiment.
6. Repeat

NLP Techniques:

Machine Learning - Sentiment Analysis

In order to figure out, if a sentence has a specific connotation, I decide to train a model to help me predict that. To train a model, you must have data to work with and be able to learn from. I used the twitter samples from the NLTK library. I retrieved the separated data, did some preprocessing on the tweets, and combined them to form one dataset. Each training example had a label attached to it. I then just simply used the NLTK Naïve Bayes Classifier to train the model. This model was then used to predict the sentiment of the sentence later on in the program. The reason for doing it so early is because it took a long time to train, so this delayed the conversation. However, if it is done earlier before the conversation starts, then the user would be able to converse seamlessly.

POS Tagging – Determine if it is a question

In order to understand the user's inputs, my program first determines, if it is a question. The way I did that, is by using the library spacy. I did some POS tagging to determine the structure of the sentence. Most questions have a format where a word with the tag W+ is followed by V+. For example, what (WP) is (VBZ) the time. So, my program determines if the input has the following structure (W+V+) by the POS tag using spacy. If so, then it is a question. If not, then the program continues to check if it is a question by using ML.

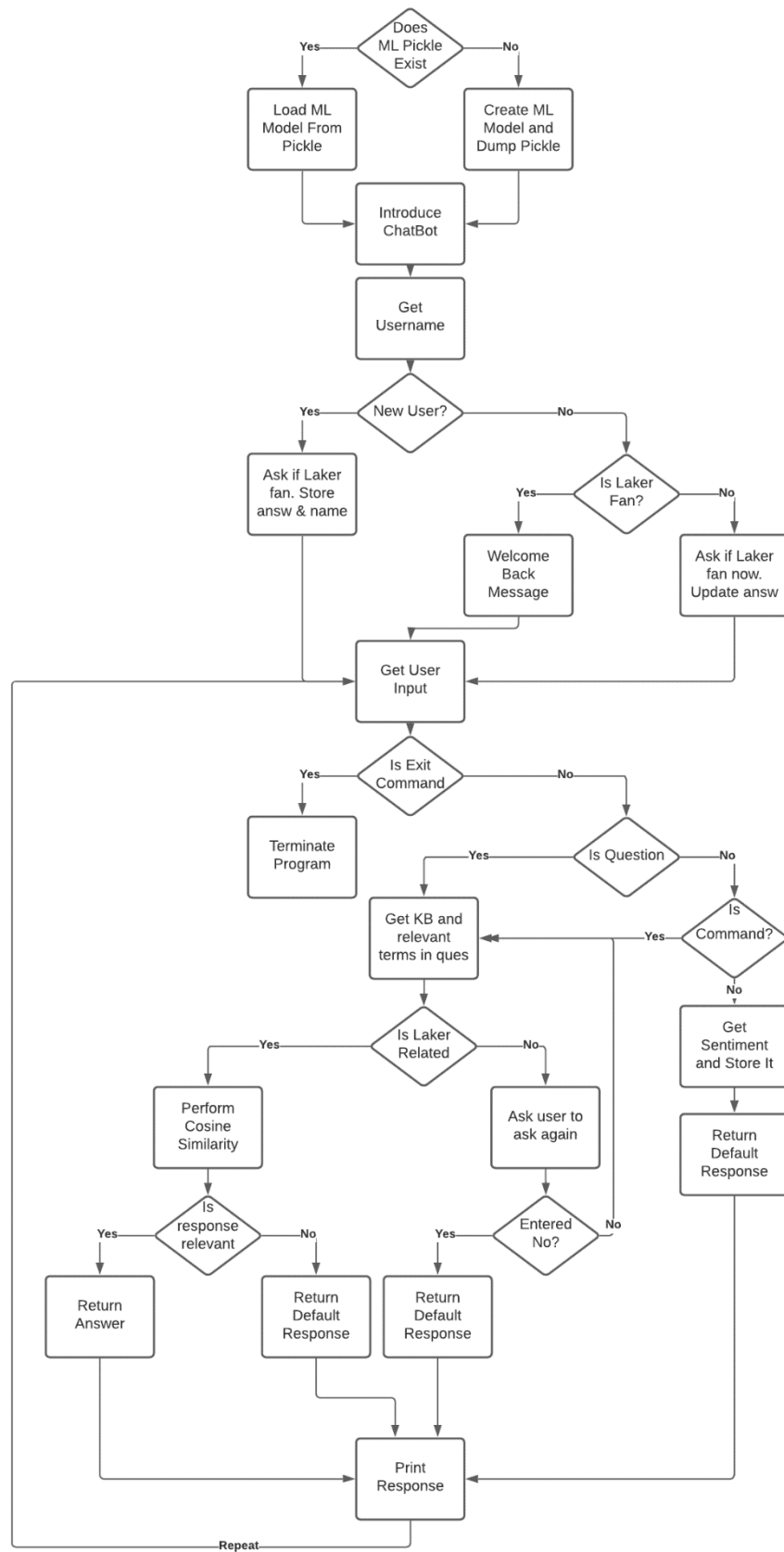
Machine Learning with Count Vectorizer – Determine if it is a question

If the POS tagging cannot determine if it is a question, then a machine learning model is used to determine if the sentence is a question. There are two files I have created to train my model. One file is filled with different types of questions and the other is filled with just statements. So, I first read the questions and statements into a data frame using pandas. I understand this is going to be a binary classification either question or not, so I added a label to each question and statement. One for question (positive) and zero for the statements (negative). I then combine the data frames to form one data frame to use for the training model. I then extract the features (sentences) and label from the data frame to use separately. I then use a count vectorizer that is fitted to the features. This will allow us to then transform our features into vectors and use it a machine learning algorithm. So that is what I do. I transform the features and the user input. With the training count vectors, I then train a logistic regression model using scikit-learn. Finally, I predict the classification of the user input and output the appropriate response.

Cosine Similarity with Tokenization & TF-IDF – Getting an answer to the question

The way my program answers question is to find the closest sentence from the knowledge base to the question. In order to do this, I must perform cosine similarity. The first thing I do is tokenize my knowledge base and the user input into sentence tokens. Each sentence is going to act like its own document. The next thing I do, is I perform TF-IDF to all the sentences or "documents" to only get the relevant terms of the knowledge base. I then fit and transform the knowledge base to this TF-IDF vectorizer from scikit-learn to transform the sentences and user input into vectors. Finally, I using scikit's learn cosine similarity, I am able to get the vector that is closest to the question.

Flow Chart:



Sample Dialog Interactions

Case 1:

```
Waking up chatbot (Training ML Model)...This might take a few minutes...
LET'S GO LAKERS.
My name is LA-BOT or call me LT for short. I am a chatbot that talks about the Los Angeles Lakers. If you tell something that is not about the Lakers I might just say some
random stuff. Also keep in mind, I am just a new born so forgive me for not understanding you at times. However, you can simply just type anything from questions to remarks,
I'll try my best to understand you.
Tell me what is your name?
Alfredo
So, Alfredo are you a fan of the Los Angeles Lakers? yes
AHHH YEAHH!!
But anyhow, let's do it! Ask or tell me anything regarding the Lakers. Type exit, bye, quit or goodbye to end the convo.
I like the lakers.
Awesome! Tell me more.
Well I grew up in LA so I that's my team.
I keep track of that.
So how are you?
Are you sure this is Laker related? If so, please ask the question one more time in a simpler way. No it's not
No worries. then. let's continue.
When did the lakers move?
Sorry. i don't know how to answer that.
Who is the owner of the Lakers
Primary lakers owner jeanie buss, who took over the team in 2017, would also become the first female controlling owner of an nba team to win the nba finals as well.
Tell me who is kobe bryant?
= 1996-2016: the kobe bryant era = 1996-2004: o'neal and bryant dynasty during the 1996 off-season, the lakers acquired 17-year-old kobe bryant from the charlotte hornets
for vlade divac; bryant was drafted 13th overall out of lower merion high school in ardmore, pennsylvania in that year's draft, by charlotte.
Who are the Los Angeles Lakers?
The los angeles lakers are an american professional basketball team based in los angeles.
Dude this is great! Thank you
Awesome! Tell me more.
No its okay
Awesome! Tell me more.
I don't want to
I don't know what to say.
Bye
PEACE!
```

Process finished with exit code 0

Case 2:

```
ONE TWO THREE ACTION!
My name is LA-BOT or call me LT for short. I am a chatbot that talks about the Los Angeles Lakers. If you tell something that is not about the Lakers I might just say some
random stuff. Also keep in mind, I am just a new born so forgive me for not understanding you at times. However, you can simply just type anything from questions to remarks,
I'll try my best to understand you.
Tell me what is your name?
Bob
So, Bob are you a fan of the Los Angeles Lakers? no
You can't be serious.
But anyhow, let's do it! Ask or tell me anything regarding the Lakers. Type exit, bye, quit or goodbye to end the convo.
Tell me who did the lakers draft in 2017
The lakers again earned the second overall pick, this time, in the 2017 nba draft, and selected lonzo ball from ucla.
how many championships do the lakers have?
The two teams have won the two highest numbers of championships, the celtics 17, the lakers 17; together, the 34 championships account for almost half of the 73 championships
in nba history.
What is the longest streak of the lakers?
By winning 33 straight games, los angeles set a record for longest winning streak of any team in american professional sports.
Who did the lakers play against in the 2020 nba finals?
They wound up beating the heat 4-2 to win the 2020 nba finals, with lebron james being named as finals mvp for the fourth time in his career.
How did the lakers get Wilt Chamberlain?
= 1968-1973: wilt arrives = on july 9, 1968, the team acquired wilt chamberlain from the philadelphia 76ers for darrell imhoff, archie clark, and jerry chambers.
I don't know what to ask anymore
Are you sure this is Laker related? If so, please ask the question one more time in a simpler way. no
No worries then. let's continue.
What does showtime mean for the lakers?
The 1980s lakers were nicknamed "showtime" due to their fast break-offense led by magic johnson.
Who is Jerry West?
The team selected jerry west from west virginia university.
okay well goodbye
Peace. PURPLE AND GOLD!

Process finished with exit code 0
```

Case 3:

```
Hey.
My name is LA-BOT or call me LT for short. I am a chatbot that talks about the Los Angeles Lakers. If you tell something that is not about the Lakers I might just say some
random stuff. Also keep in mind, I am just a new born so forgive me for not understanding you at times. However, you can simply just type anything from questions to remarks,
I'll try my best to understand you.
Tell me what is your name?
Bob
Bob!!! Welcome back. Glad you're here.
Have you changed your mind of becoming a Lakers fan? yes
AHHH YEAHH!!
But anyhow, let's do it! Ask or tell me anything regarding the Lakers. Type exit, bye, quit or goodbye to end the convo.
what are the colors of the lakers?
The team's colors are purple, gold and white.
Who are the rivals of the lakers?
The lakers won two of the first three games.
That wasn't the answer I was expecting
I am sorry you feel that way.
This thing is not that very smart
I appreciate your honesty.
Can you tell me something about Magic Johnson
Sorry. I can figure out if it's a question or not. Is it? yes
The 1980s lakers were nicknamed "showtime" due to their fast break-offense led by magic johnson.
Are you like siri?
Sorry. I can figure out if it's a question or not. Is it? yes
yes
Are you sure this is Laker related? If so, please ask the question one more time in a simpler way. no
No worries then. let's continue.
exit
Bye.

Process finished with exit code 0
```

Appendix for Knowledge Base

The knowledge base used is from the Wikipedia API. What the program does, is that it retrieves the Lakers Wikipedia page and sentence tokenize the content. Then cosine similarity is performed on each sentence with the user input to get the response. The knowledge base is shown below.

Note: The knowledge base shown is below is NOT the complete knowledge base used in the program. The whole KB is not placed in here because it would take too much space.

['the los angeles lakers are an american professional basketball team based in los angeles.', 'the lakers compete in the national basketball association (nba) as a member of the league's western conference pacific division.', 'the lakers play their home games at staples center, an arena shared with the nba's los angeles clippers, the los angeles sparks of the women's national basketball association, and the los angeles kings of the national hockey league.', 'the lakers are one of the most successful teams in the history of the nba, and have won 17 nba championships, tied with the boston celtics for the most in nba history.the franchise began with the 1947 purchase of a disbanded team, the detroit gems of the national basketball league (nbl).', 'the new team began playing in minneapolis, minnesota calling themselves the minneapolis lakers.', 'initially a member of the nbl, the lakers won the 1948 nbl championship before joining the rival basketball association of america, where they would win five of the next six championships, led by star george mikan.', 'after struggling financially in the late 1950s following mikan's retirement, they relocated to los angeles before the 1960-61 season.', 'led by hall of famers elgin baylor and jerry west, los angeles made the nba finals six times in the 1960s, but lost each series to the celtics, beginning their long and storied rivalry.', 'in 1968, the lakers acquired four-time nba most valuable player (mvp) wilt chamberlain, and won their sixth nba title-and first in los angeles-in 1972, led by new head coach bill sharman.', 'after the retirement of west and chamberlain, the team acquired kareem abdul-jabbar, who also won multiple mvp awards, but was unable to make the finals in the late 1970s.', 'the 1980s lakers were nicknamed "showtime" due to their fast break-offense led by magic johnson.', 'the team won five championships in a nine-year span, and contained hall of famers johnson, abdul-jabbar, and james worthy, and was led by hall of fame coach pat riley.', 'after abdul-jabbar and johnson retired, the team struggled in the early 1990s, before acquiring shaquille o'neal and kobe bryant in 1996.', 'with the duo, who were led by another hall of fame coach phil jackson, the team won three consecutive titles between 2000 to 2002, securing the franchise its second "three-peat".', 'led by bryant, the lakers won two more championships in 2009 and 2010, but failed to regain their former glory in the following decade.', 'led by lebron james and anthony davis, the lakers claimed their 17th championship in the 2020 nba finals.the lakers hold the record for nba's longest winning streak, 33 straight games, set during the 1971-72 season.', 'twenty-six hall of famers have played for los angeles, while four have coached the team.', 'four lakers-abdul-jabbar, johnson, o'neal, and bryant-have won the nba mvp award for a total of eight awards.', 'team history = 1947-1958: beginnings and minneapolis dynasty = the lakers' franchise began in 1947 when ben berger and morris chalfen of minnesota purchased the recently disbanded detroit gems of the national basketball league (nbl) for \$15,000 from gems owner maury winston.', 'minneapolis sportswriter sid hartman played a key behind the scenes role in helping put together the deal and later the team.', 'inspired by minnesota's nickname, "land of 10,000 lakes", the team christened themselves the lakers.', 'hartman helped them hire john kundla from college of st. thomas, to be their first head coach, by meeting with him and selling him on the team.the lakers had a solid roster, which featured forward jim pollard, playmaker herm schaefer, and center george mikan, who became the most dominant player in the nbl.', 'in their first season, they led the league with a 43-17 record, later winning the nbl championship that season.', 'in 1948, the lakers moved from the nbl to the basketball association of america (baa), and mikan's 28.3 point per game (ppg) scoring average set a baa record.', 'in the 1949 baa finals they won the championship, beating the washington capitols four games to two.', 'the following season, the team improved to 51-17, repeating as champions.', 'in the 1950-51 season, mikan won his third straight scoring title at 28.4 ppg and the lakers went 44-24 to win their second straight division title.', 'one of those games, a 19-18 loss against the fort wayne pistons, became infamous as the lowest scoring game in nba history.', 'in the playoffs, they defeated the indianapolis olympians in three games but lost to the rochester royals in the next round.during the 1951-52 season, the lakers won 40 games, finishing second in their division.', 'they faced the new york knicks in the nba finals, which they won in seven games.', 'in the 1952-53 season, mikan led the nba in rebounding, averaging 14.4 rebounds per game (rpg), and was named mvp of the 1953 nba all-star game.', 'after a 48-22 regular season, the lakers defeated the fort wayne pistons in the western playoffs to advance to the nba finals.', 'they then defeated the new york knicks to win their second straight championship.', 'though lakers star george mikan suffered from knee problems throughout the 1953-54 season, he was still able to average 18 ppg.', 'clyde lovellette, who was drafted in 1952, helped the team win the western division.', 'the team won its third straight championship in the 1950s and fifth in six seasons when it defeated the syracuse nationals in seven games.following mikan's retirement in the 1954 off-season, the lakers struggled but still managed to win 40 games.', 'although they defeated the rochester royals in the first round of the playoffs, they were defeated by the fort wayne pistons in the semifinals.', 'although they had losing records the next two seasons, they made the playoffs each year.', 'mikan came back for the last half of the 1955-56 season, but struggled and retired for good after the season.', 'led by lovellette's 20.6 points and 13.5 rebounds, they advanced to the conference finals in 1956-57.', 'the lakers had one of the worst seasons in team history in 1957-58 when they won a league-low 19 games.', 'they had hired mikan, who had been the team's general manager for the previous two seasons, as head coach to replace kundla.', 'mikan was fired in january when the team was 9-30, and kundla was rehired.the lakers earned the top pick in the 1958 nba draft and used it to select elgin baylor.', 'baylor, who was named nba rookie of the year and co-mvp of the 1959 nba all-star game, averaged 24.9 ppg and 15.0 rpg helping the lakers improve to second in their division despite a 33-39 record.', 'after upsetting the hawks in six games in the division finals, they returned to the nba finals, but were swept by the celtics, beginning their long rivalry.', '= 1958-1968: move to los angeles and celtics rivalry = in their last year in minneapolis, the lakers went 25-50.', 'on january 18, 1960, the team was coming off a loss and traveling to st. louis when their plane crash-landed.', 'snow storms had driven the pilot 150 miles off course when he was forced to land in a cornfield.', 'no one was hurt.', 'their record earned them the number two pick in the 1960 nba draft.', 'the team selected jerry west from west virginia university.', 'during the 1960 off-season, the lakers became the nba's first west coast team when owner bob short decided to move the team to los angeles.', 'led by baylor's 34.8 ppg and 19.8 rpg, los angeles won 11 more than the year before in west's first season.', 'on november 15 that season, baylor set a new nba scoring record when he scored 71 points in a victory against the new york knicks while grabbing 25 rebounds.', 'in doing so, baylor broke his own nba record of 64 points.', 'despite a losing record, the lakers made the playoffs.', 'they came within two points of the nba finals when they lost in game seven of their second round series against st. louis.led by baylor and west at 38.3 and 30.8 ppg respectively, the lakers improved to 54-26 in 1961-62, and made the finals.', 'in a game five victory, baylor grabbed 22 rebounds and set the still-standing nba record for points in a finals game with 61, despite fouling out of the game.', 'the lakers, however, lost to the celtics by three points in overtime of game seven.', 'frank selvy, after making two jumpers in the final 40 seconds to tie the game, missed a potential game-winning 18 foot jump shot in regulation, a miss which he said in june 2010 still haunted him more than 40 years later.los angeles won 53 games in 1962-63, behind baylor's 34.0 ppg and west's 27.1 ppg but lost in the nba finals in six games to the celtics.', 'after falling to 42-38 and losing in the first round of the 1964 nba playoffs to the hawks, the team won 49 games in 1964-65.', 'the lakers surged past the baltimore bullets in the division finals, behind west's record-setting 46.3 ppg in the series.', 'they lost again to celtics in the finals however, this time in five games.los angeles lost in the finals to boston in seven games again in 1966, this time by two points.', 'down by 16 entering the fourth quarter, and 10 with a minute and a half to go, the lakers mounted a furious rally in the closing moments, which fell just short.', 'after dropping to 36 wins and losing in the first round of the 1967 nba playoffs, they lost in the finals to the celtics again in 1968.', 'los angeles moved to a brand-new arena, the forum, in 1967, after playing seven seasons at the los angeles memorial sports arena.', '= 1968-1973: wilt arrives = on july 9, 1968, the team acquired wilt chamberlain from the philadelphia 76ers for darrell imhoff, archie clark, and jerry chambers.', 'in his first season as a laker, chamberlain set a team record by averaging a league-leading 21.1 rpg.', 'west, baylor, and chamberlain each averaged over 20 points, and los angeles won their division.', 'the lakers and celtics again met in the finals, and los angeles had home court advantage against boston for the first time in their rivalry.', 'they won the first game behind jerry west's 53 points, and had a 3-2 lead after five.', 'boston won the series in seven games however, and earned their 11th nba championship in 13 seasons.', 'west was named the first-ever finals mvp; this remains the only time that a member of the losing team has won the award.', 'in 1970, west won his first scoring title at 31.2 ppg, the team returned to the finals, and

Appendix for User Models

The appendix for the user models is divided into two database files. The first database file only has one table with two columns, one for the username which is the primary key, and the other one to that simply says “yes” or “no” to the question if the user is a Laker fan. This information is used in the introduction of the chatbot. The other database file could potentially have many tables. Each table is named after the username that is currently in session with the chatbot. Each table has three columns. The first column is an autoincrement unique id, the second column is the statement the user has made, and the third column is simply a sentence saying “positive” or “negative” describing the sentiment of the sentence. It is difficult to open the files through windows, so I have extracted the database files in a python program and converted them to a list to print them. The following two databases are shown below.

```
[('alfredo', 'yes'), ('bob', 'yes'),  
  
[('alfredo', 'yes'), ('bob', 'yes'), ('alberto', 'yes'), ('josh', 'no'), ('george', 'yes'), ('james', 'yes'), ('maria', 'no')]  
  
[('alfredo', 'yes'), ('bob', 'yes'), ('alberto', 'yes'), ('josh', 'no'), ('george', 'yes'), ('james', 'yes'), ('maria', 'no'), ('juan', 'no'), ('rick', 'yes')]  
  
[(1, 'the lakers are my team!', 'Positive')]
```

Evaluation

Strengths:

- Can answer simple questions that involve the lakers, especially when the key words are align with the type of answer you want.
- Can handle any input and it will generate a response.
- It can sometimes tell when the question is not related to the lakers.
- It can determine if the answer is relevant or not, if not it doesn't show it.
- Use machine learning to perform sentiment analysis and predict if it is a question or not.
- Understands that some queries are not in a question format, such as “tell me about the lakers”.

Weakness:

- Does not actually understand the conversation occurring.
- Can give unexpected answers that have nothing to related to the question but because some key words were in there, it prints that out.
- Not a lot of human interaction, for example, it lacks the small talk.
- Can sometimes confuse statements as questions
- Can sometimes predict the sentiment incorrectly

Survey Results:

1. In a scale from 1 to 5, rate the answers given by the LA-BOT.
Responses: 3,2,3
2. In a scale from 1 to 5, rate the responses given by the LA-BOT.
Reponses: 3, 2, 2
3. In a scale from 1 to 5, how satisfactory are you with LA-BOT.
Responses: 4,3,3
4. In a scale from 1 to 5, rate the understanding of what you were trying to ask to LA-BOT.
Responses: 2,1,2
5. In a scale from 1 to 5, how confident are you this can be used commercially.
Responses: 1,1,1