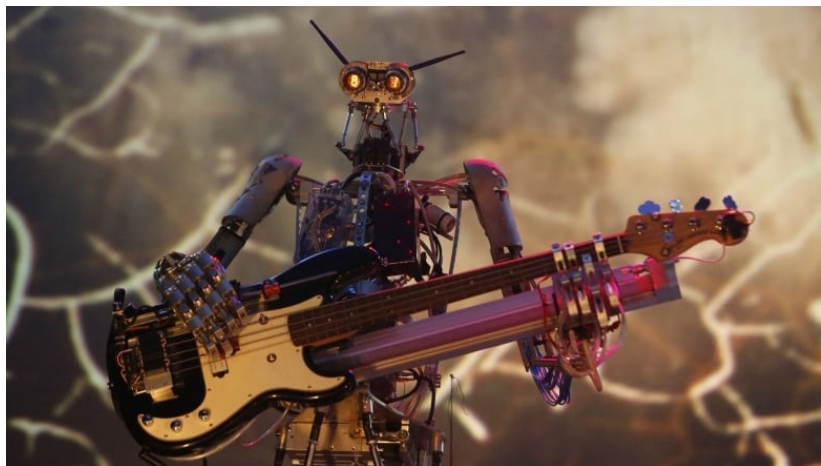




Generative Music with RNN-LSTM



AI & Music 

Rocco Zaccagnino
rzaccagnino@unisa.it

Creative AI



Creative AI

euro
news.

BEETHOVEN'S 250TH ANNIVERSARY

AI is writing the end of
unfinished symphony

BEETHOVEN ORCHESTRA BONN
CAMERON CARPENTER
DIRK KAFTAN

BEETHOVEN
X
THE AI PROJECT

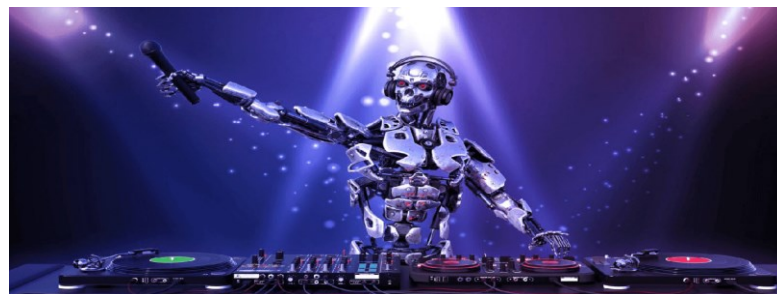
Generative Music

Forma di musica in cui un brano musicale si crea da un insieme iniziale di elementi musicali, comportamenti e regole definiti dal compositore e/o da un sistema (naturale o **artificiale**).

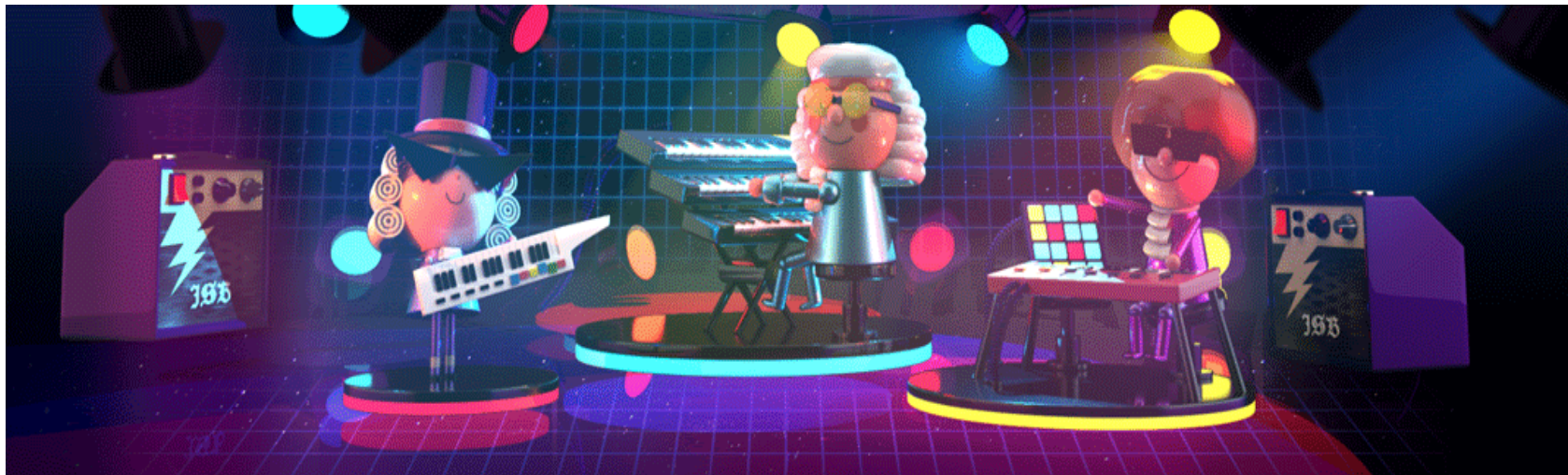
"Approach to music creation concerning itself with neither improvisation nor explicit composition, but rather with framing an indeterminate system from which music can emerge" (Priestley, 2014)



"Generative music is like trying to create a seed, as opposed to classical composition which is like trying to engineer a tree" (Toop, 2004)



Generative Music



Non è un genere musicale o uno stile, ma una pratica compositiva in cui il compositore è più interessato a creare un sistema fisico o virtuale che poi genererà la musica autonomamente

Music & NN

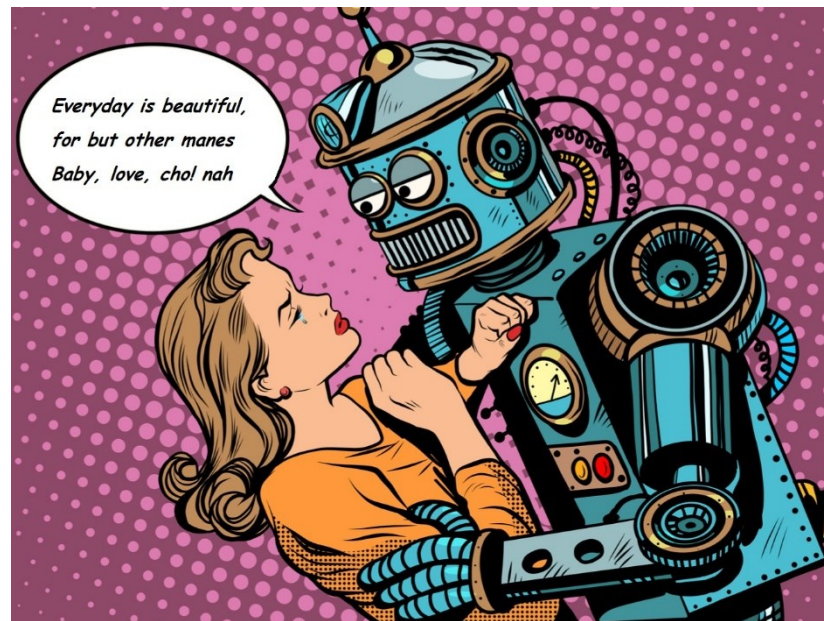
Obiettivo: sviluppare una RNN-LSTM per generare **melodie** automaticamente



- Build e train di una **LSTM** in Keras
- Gestire **time-series** data
- Comprensione **rappresentazione musicale simbolica**
- Concetti musicali di base (e.g., pitch, duration, key)
- Pre-processing di musica simbolica

Pre-requisiti & Tools

- Intermediate Python
- Familiarità con **Tensorflow** & **Keras**
- **Music21** (MIT music library)
- **Musescore** notation platform
- Ubuntu operating system (v. 22.04)



"Artificial intelligence just wrote the best love song of all time (and it's terrible)" **Verdict**

Melody

Successione di suoni animata dal ritmo e regolata da leggi strofiche così da acquistare contorni, e senso propri ... parte di composizione musicale che costituisce il canto, il motivo fondamentale, come successione lineare di suoni che, accompagnata da altri suoni secondo regole determinate, è la base dell'armonia.

Treccani, <https://www.treccani.it/vocabolario/melodia/>

Da un punto di vista applicativo ... sequenza di note e pause



Melody: pitch

Indica quanto “alta” o “bassa” è la nota. Detta anche **frequenza**.



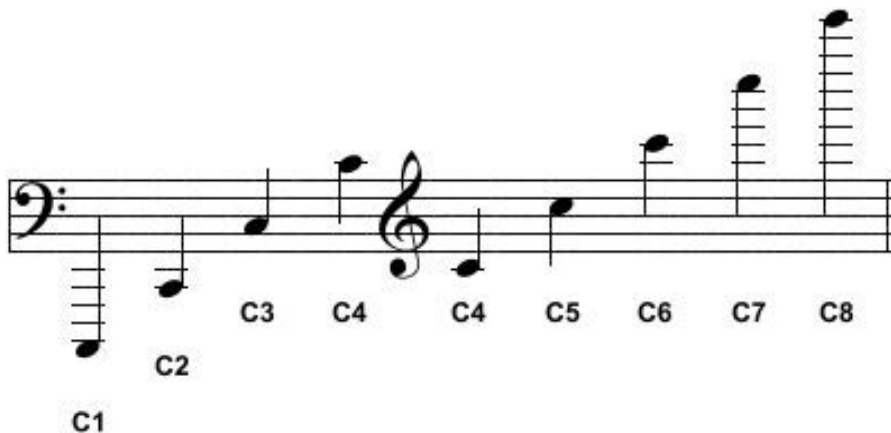
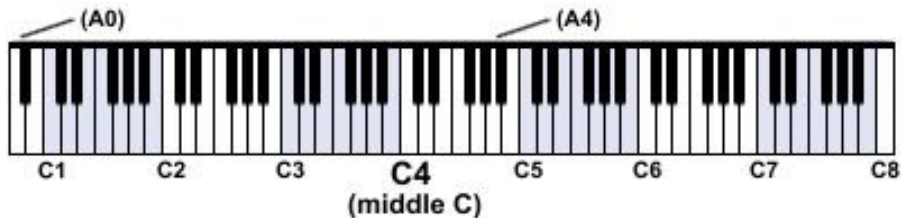
Melody: pitch

Indica quanto “alta” o “bassa” è la nota. Detta anche **frequenza**.



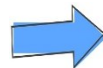
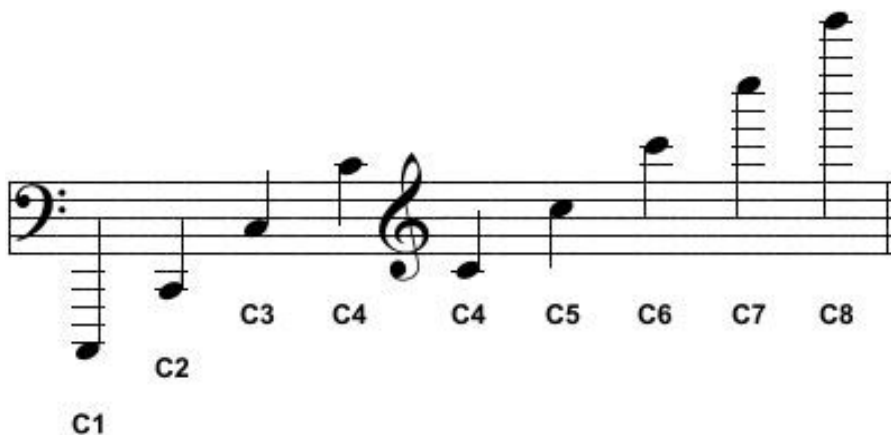
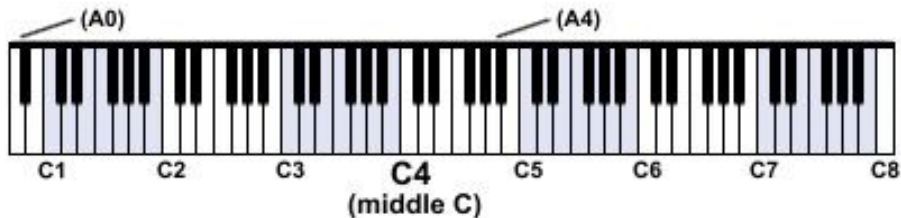
Scientific pitch notation

Note name + octave (e.g., C3, D5,...)



Scientific pitch notation

Note name + octave (e.g., C3, D5,...)



MIDI note notation

MIDI è uno standard per suonare, editare e registrare musica

Fornisce un **mapping** di nomi di note a interi (**C4 : 60**)

MIDI note notation

MIDI è uno standard per suonare, editare e registrare musica

Fornisce un **mapping** di nomi di note a interi (**C4 : 60**)

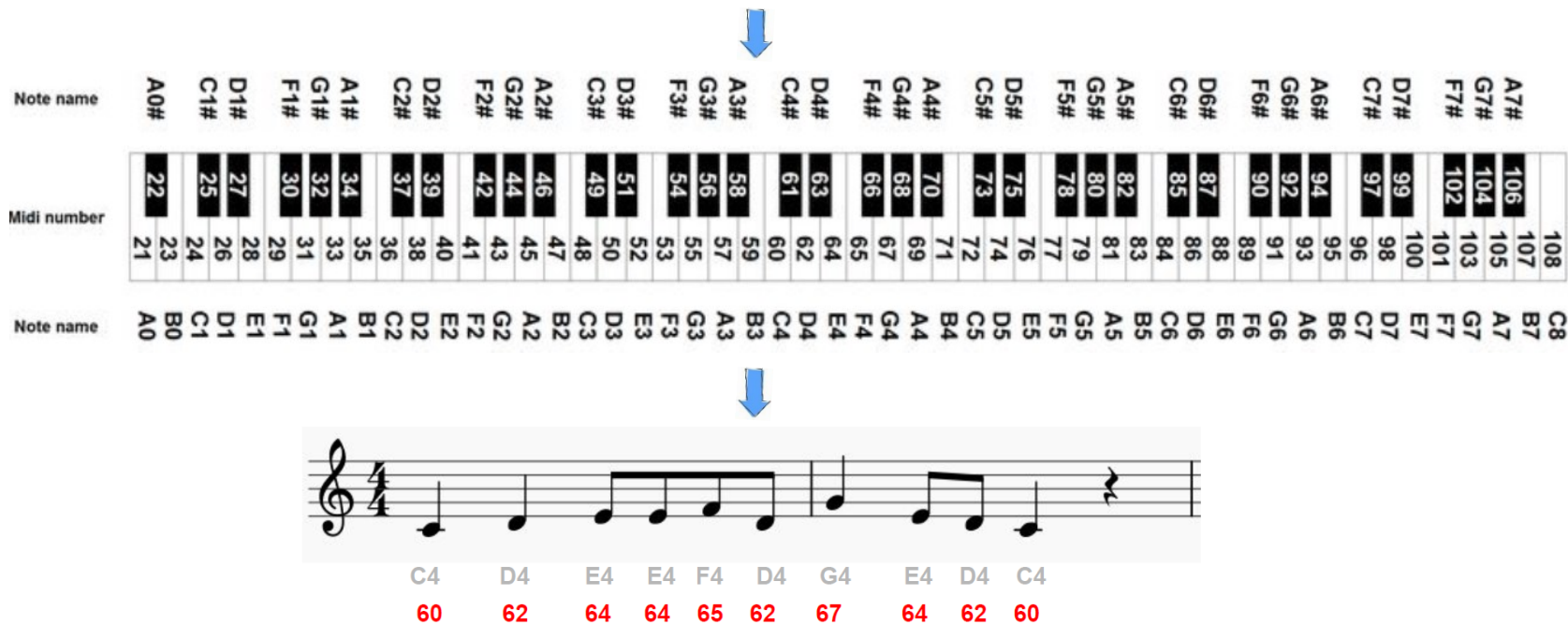


A7#	106	107	C8
G7#	104	105	A7
F7#	102	103	G7
D7#	99	100	F7
C7#	97	98	E7
A6#	94	95	D7
G6#	92	93	C7
F6#	90	91	B6
D6#	87	88	A6
C6#	85	86	G6
A5#	82	83	F6
G5#	80	81	E6
F5#	78	79	D6
D5#	75	76	C6
C5#	73	74	B5
A4#	70	71	A5
G4#	68	69	G5
F4#	66	67	F5
D4#	63	64	E5
C4#	61	62	D5
A3#	58	59	C5
G3#	56	57	B4
F3#	54	55	A4
D3#	51	52	G4
C3#	49	50	F4
A2#	46	47	E4
G2#	44	45	D4
F2#	42	43	C4
D2#	39	40	B3
C2#	37	38	A3
A1#	34	35	G3
G1#	32	33	F3
F1#	30	31	E3
D1#	27	28	D3
C1#	25	26	C3
A0#	22	23	B2
		21	A2

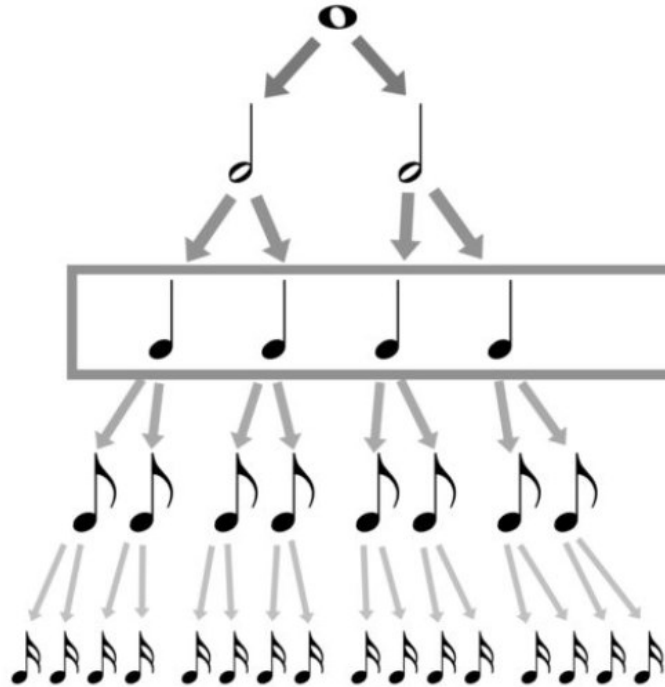
MIDI note notation

MIDI è uno standard per suonare, editare e registrare musica

Fornisce un **mapping** di nomi di note a interi (**C4 : 60**)



Note durations



1 whole note = 4 beats

1 half note = 2 beats

1 quarter note = 1 beat

1 eighth note = $\frac{1}{2}$ a beat

1 sixteenth note = $\frac{1}{4}$ a beat

Note durations



Duration in # beats:

4

4

Time signature

beats in a bar




Type of note which equals 1 beat


Time Signature

Beat Duration


Number of Beats

$\frac{3}{2}$

$\frac{2}{2}$ = 


$\frac{3}{2}$ = 

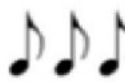
$\frac{3}{4}$

$\frac{4}{4}$ = 


$\frac{3}{4}$ = 


$\frac{3}{8}$

$\frac{8}{8}$ = 

$\frac{3}{8}$ = 

$\frac{3}{16}$

$\frac{16}{16}$ = 

$\frac{3}{16}$ = 

Music representation 1

Sequence

(pitch, duration) per ogni nota

E.g., [(C4,1),(D4,1),(E4,.5),...]



A musical score in 4/4 time, treble clef, showing a sequence of notes. The notes are: C4 (quarter), D4 (quarter), E4 (quarter), E4 (quarter), F4 (quarter), D4 (quarter), G4 (quarter), E4 (quarter), D4 (quarter), C4 (quarter), and a whole rest. The duration for each note is 1, except for the whole rest which is 2. The pitch and duration are listed below each note in red text.

Note	Pitch	Duration
C4	C4	1
D4	D4	1
E4	E4	.5
E4	E4	.5
F4	F4	.5
D4	D4	.5
G4	G4	1
E4	E4	.5
D4	D4	.5
C4	C4	.5
Rest	r	1

Music representation 2

Time series

Campionamento della melodia ad ogni 16th note

Step: 16th note

Log MIDI note quando occorre la nota

Usiamo il simbolo “_” per indicare la **nota tenuta**

Usiamo il simbolo “r” per indicare una **pausa**

- 4/4 time signature
- 16 samples per bar
- 4 samples per quarter note



Music representation 2 (example)

["60", "_", "_", "_",



Music representation 2 (example)

["60", "_", "_", "_",

"62", "_", "_", "_",



Music representation 2 (example)

["60", "_", "_", "_",

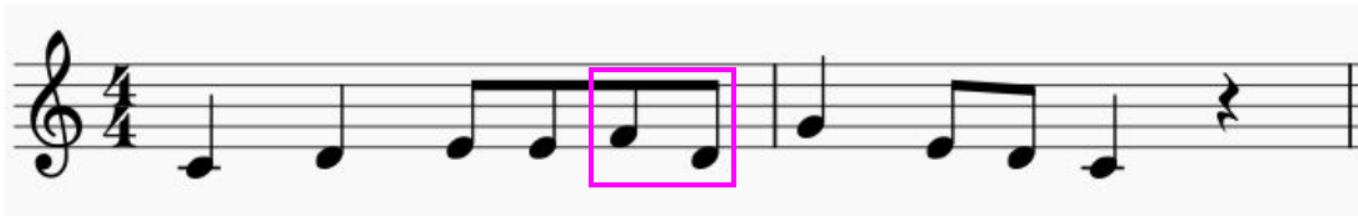
"62", "_", "_", "_",

"64", "_", "64", "_",



Music representation 2 (example)

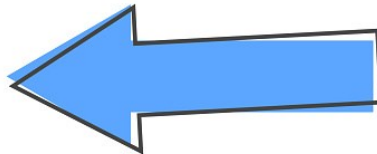
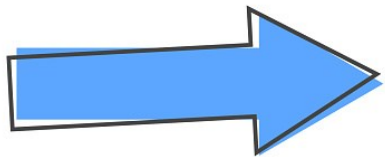
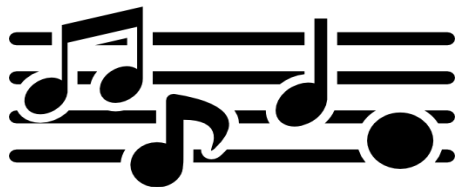
["60", "_", "_", "_",
"62", "_", "_", "_",
"64", "_", "64", "_",
"65", "_", "62", "_",
...]



Melody generation problem

Considerare la melodia come “time-series data”

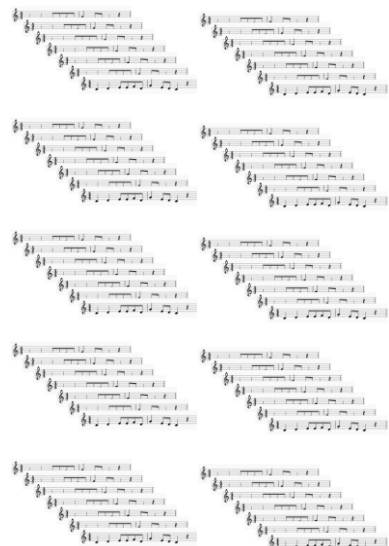
Melody generation
problem



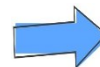
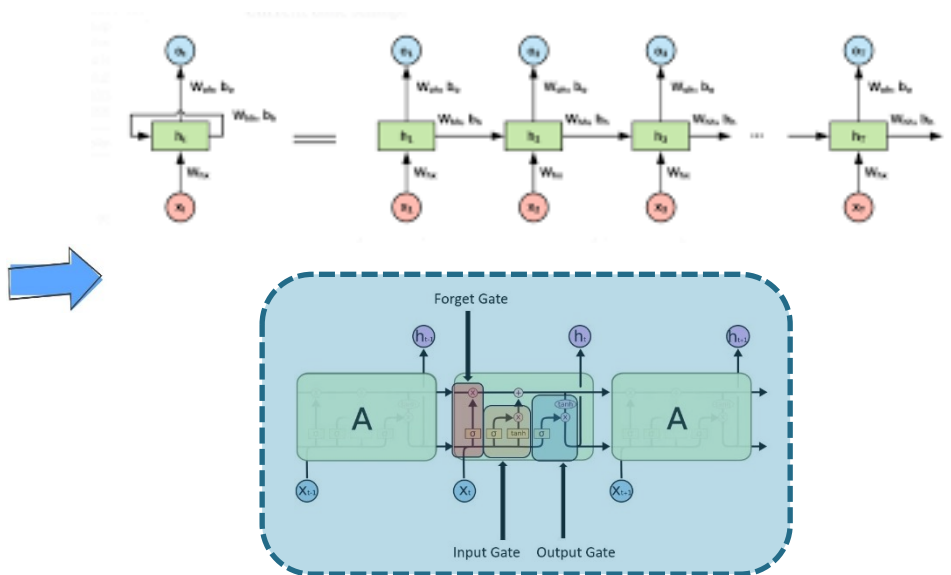
Time-series prediction
problem

The melody generator (training)

Melodies dataset



RNN-LSTM



Next note
prediction

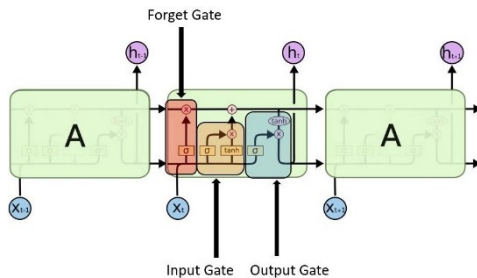
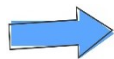
The melody generator (inference)

1



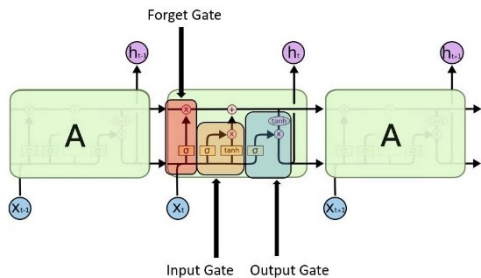
The melody generator (inference)

2



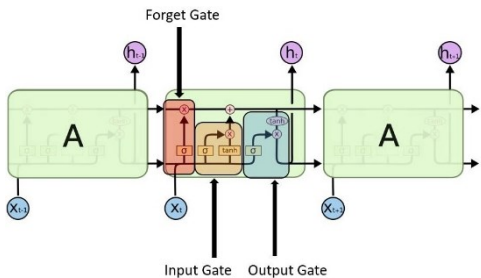
The melody generator (inference)

3



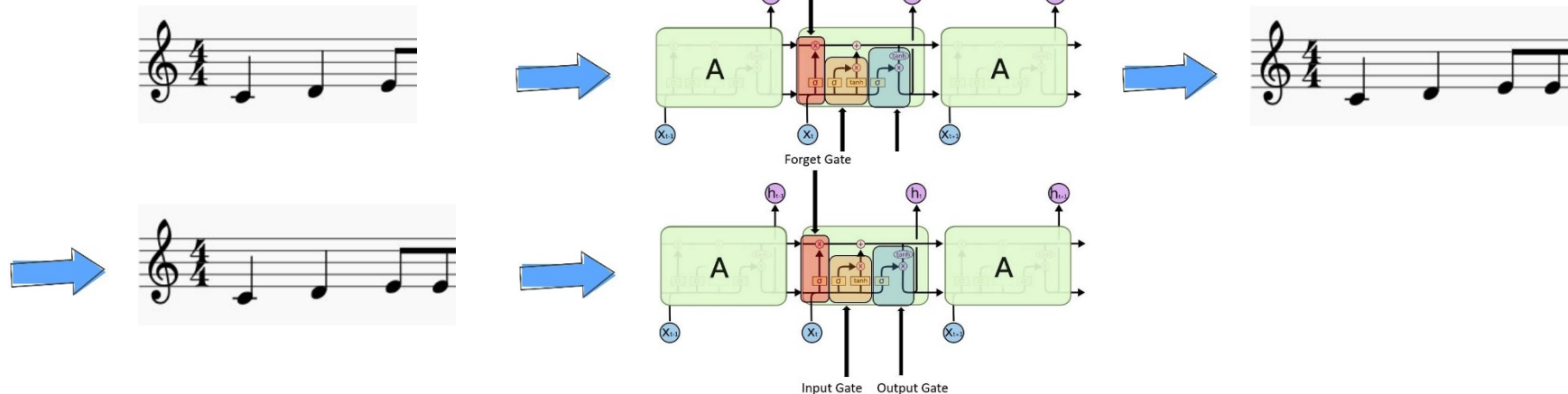
The melody generator (inference)

4



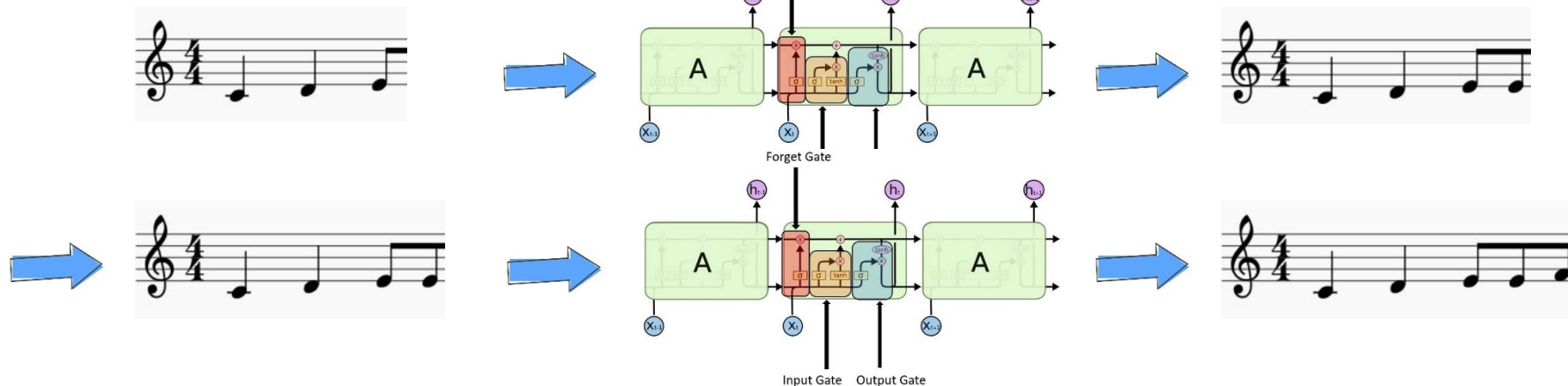
The melody generator (inference)

5



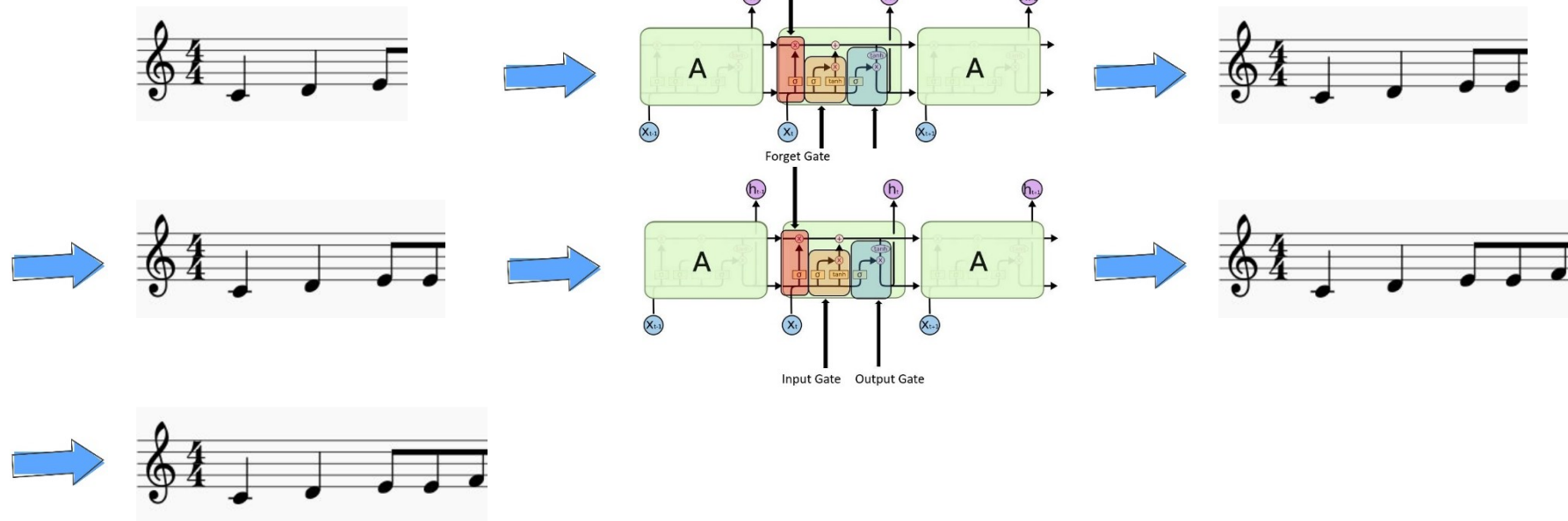
The melody generator (inference)

6



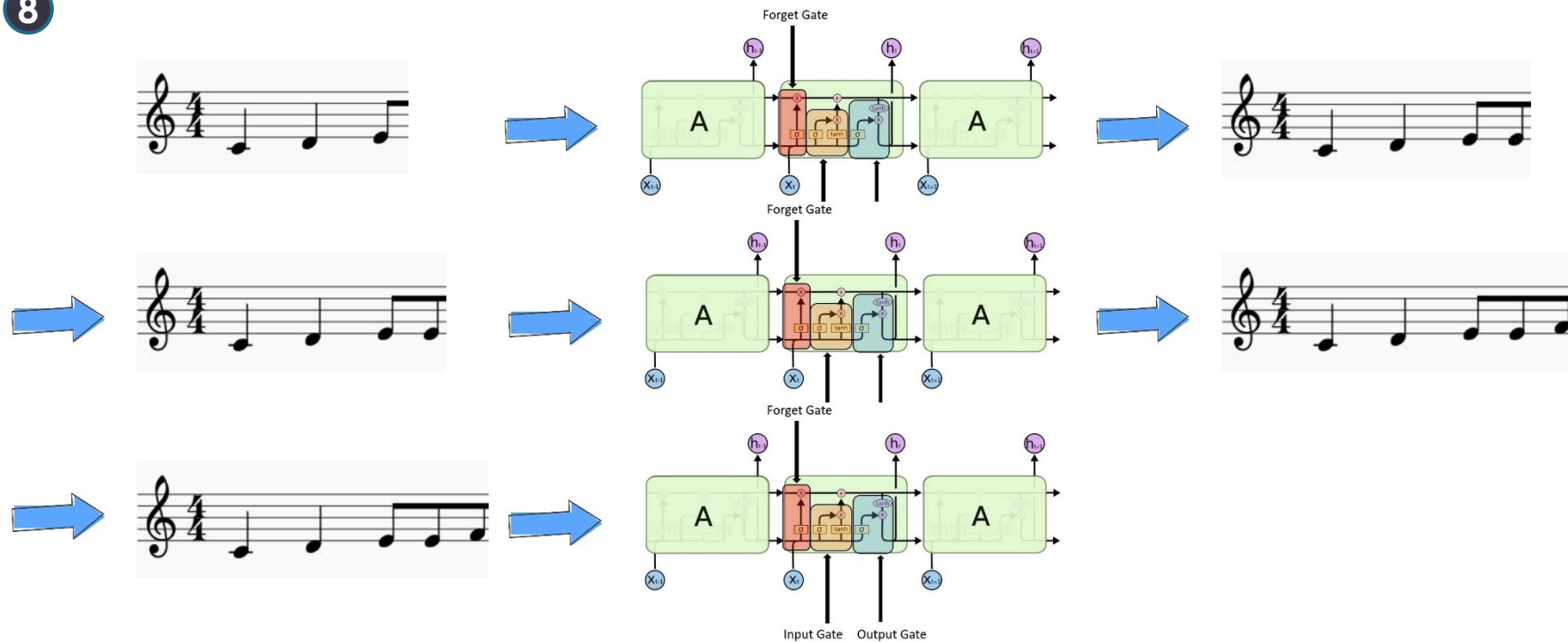
The melody generator (inference)

7



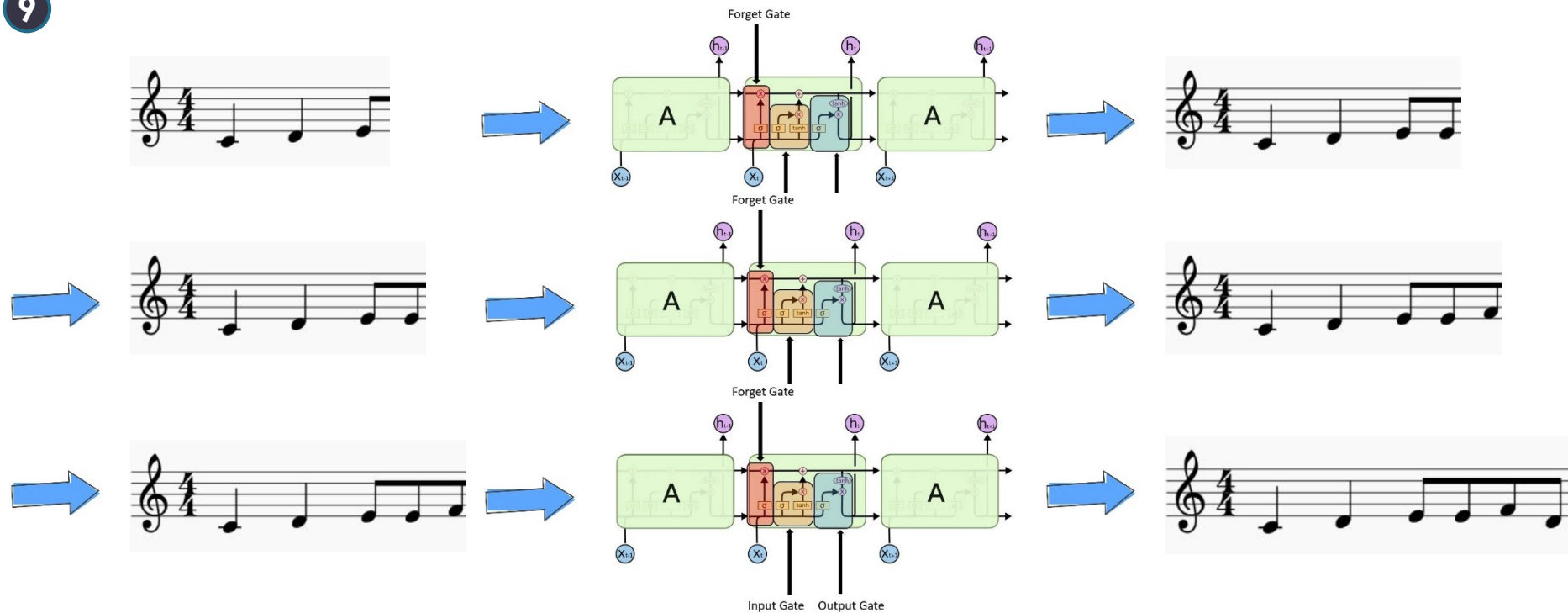
The melody generator (inference)

8



The melody generator (inference)

9

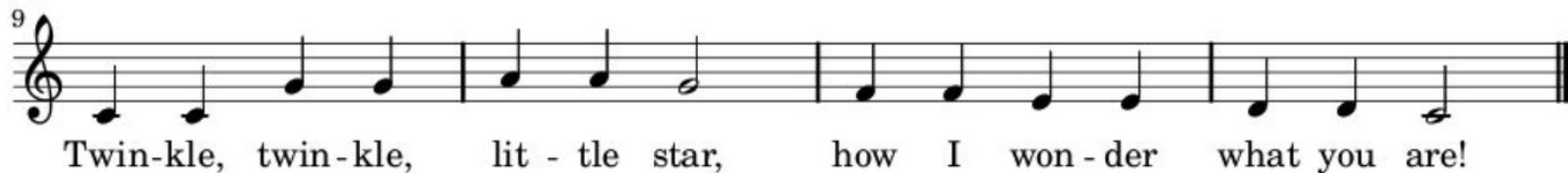
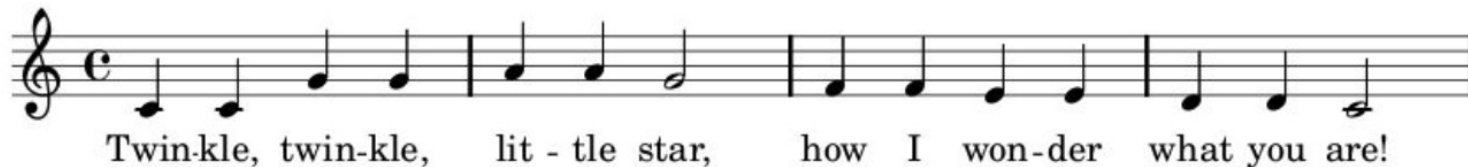


Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*

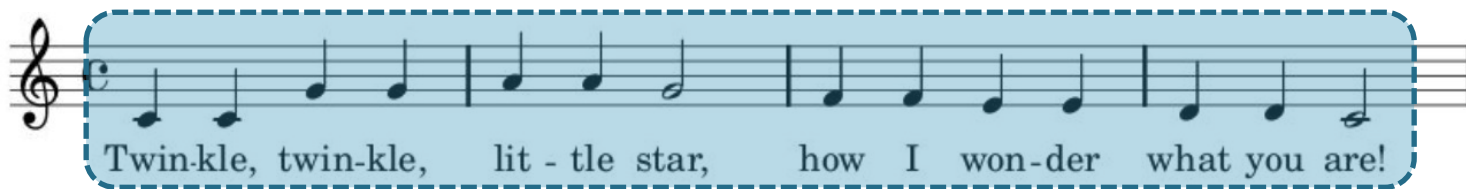
Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*



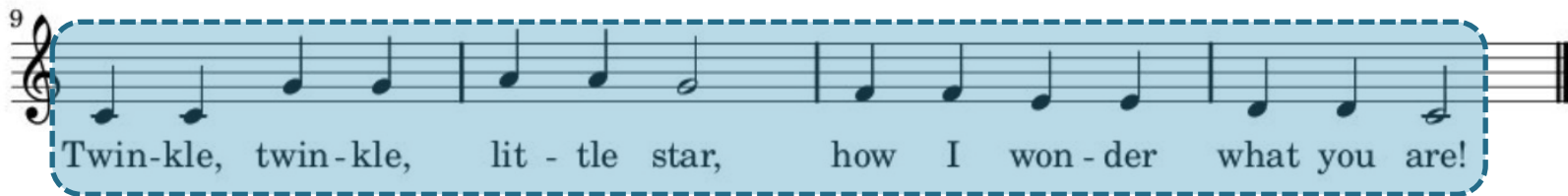
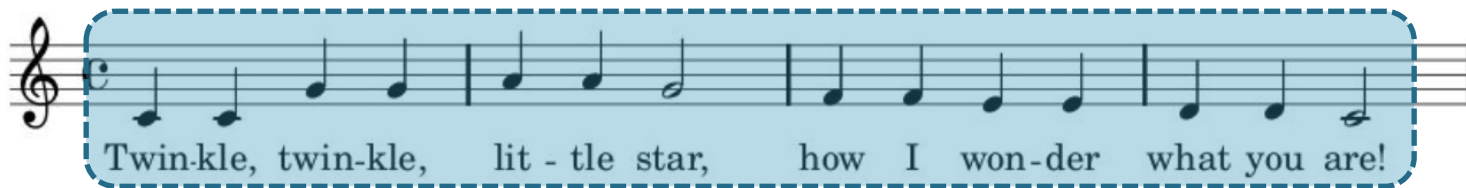
Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*



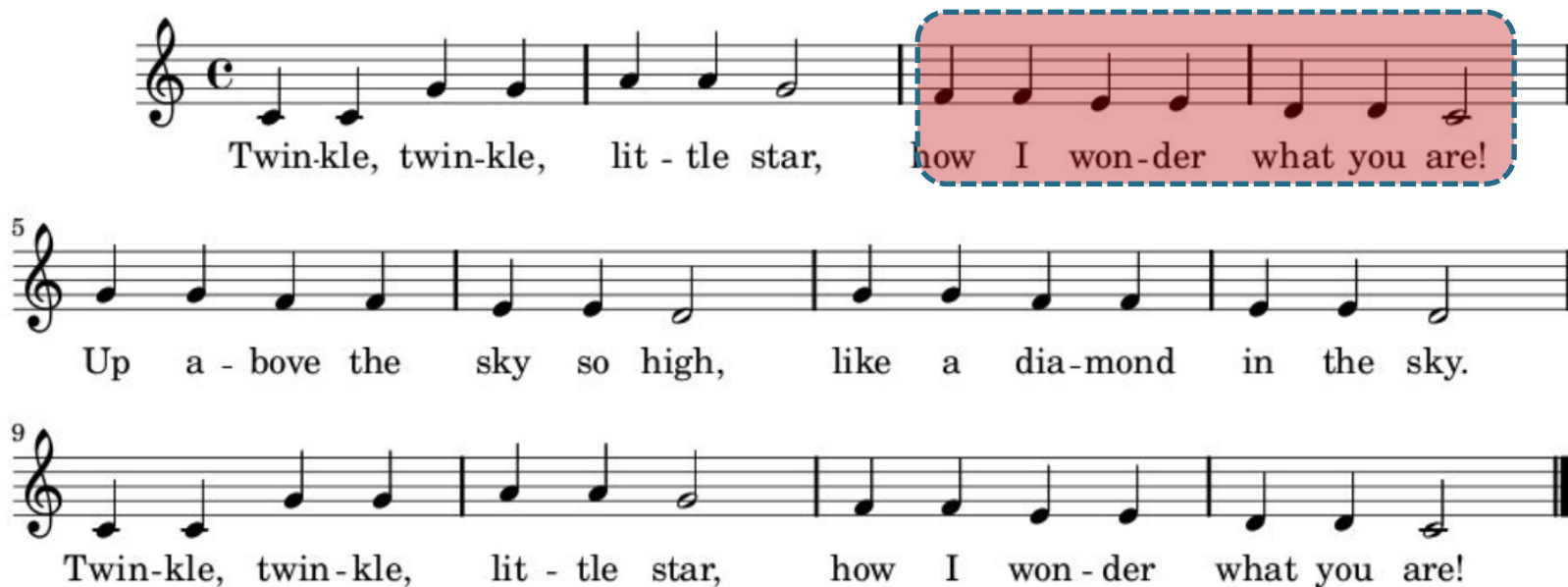
Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*



Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*



The image displays a musical score for the song "Twinkle, Twinkle, Little Star" in treble clef with a common time signature (C). The score is divided into three systems. The first system contains the first line of music, with the lyrics "Twinkle, twinkle, lit - tle star," followed by a red dashed box highlighting the melody for "how I won-der what you are!". The second system starts at measure 5 and contains the lyrics "Up a - bove the sky so high, like a dia-mond in the sky." The third system starts at measure 9 and contains the lyrics "Twinkle, twinkle, lit - tle star, how I won - der what you are!". The melody in the first system is a simple, repetitive pattern of eighth and quarter notes, which is repeated in the third system, illustrating long-term structural patterns.

Twinkle, twinkle, lit - tle star, how I won-der what you are!

5 Up a - bove the sky so high, like a dia-mond in the sky.

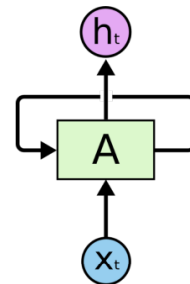
9 Twinkle, twinkle, lit - tle star, how I won - der what you are!

Perché utilizzare RNN-LSTM?

Le melodie hanno *long-term structural patterns*

Le LSTMs catturano *long-term structural dependencies*

Meccanismo di ricorrenza: **RNN**



FNN:

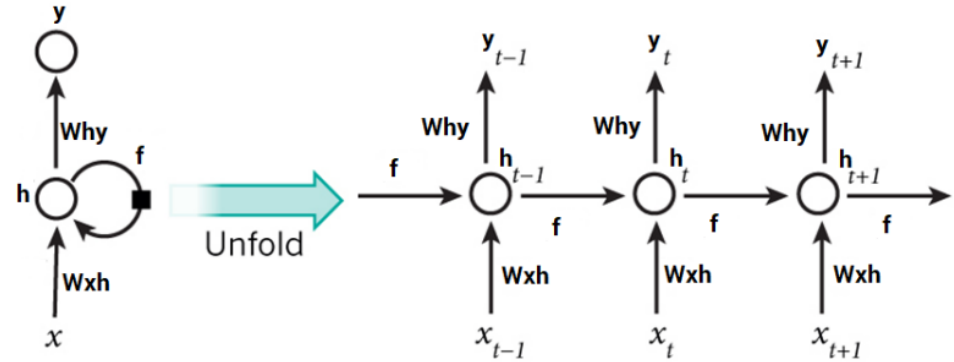
1. **No loops:** informazione in una sola direzione
2. **No memory:** decisioni in base all'input corrente
3. **No series:** non gestiscono time-series data

RNN:

1. **Gestiscono time-series data**
2. **Considerano input corrente e precedente**
3. **Hanno un meccanismo di memoria**

Recurrent Neural Network: RNN

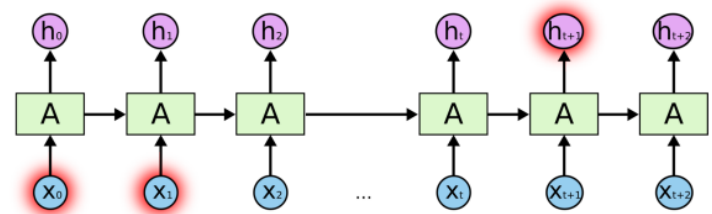
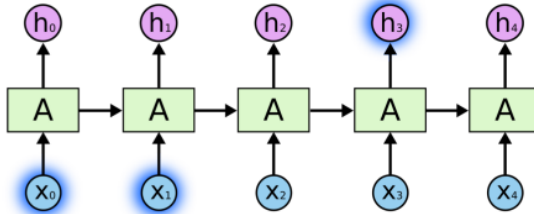
$$h_t = f(W_{xh} x_t + W_{hy} h_{t-1})$$



- h_t : hidden state at time step t
- x_t : input at time step t
- W_{xh} and W_{hy} : weight matrices. Filters that determine how much importance to accord to both the present input and the past hidden state.

Recurrent Neural Network: RNN

- A small example where RNN can work perfectly :
 - Prediction of the last word in the sentence : “The clouds are in the sky”
- RNN can't handle situation where the **gap** between the **relevant information** and the point where it is needed is **very large**.



Long-Short Term Memory Network: LSTM

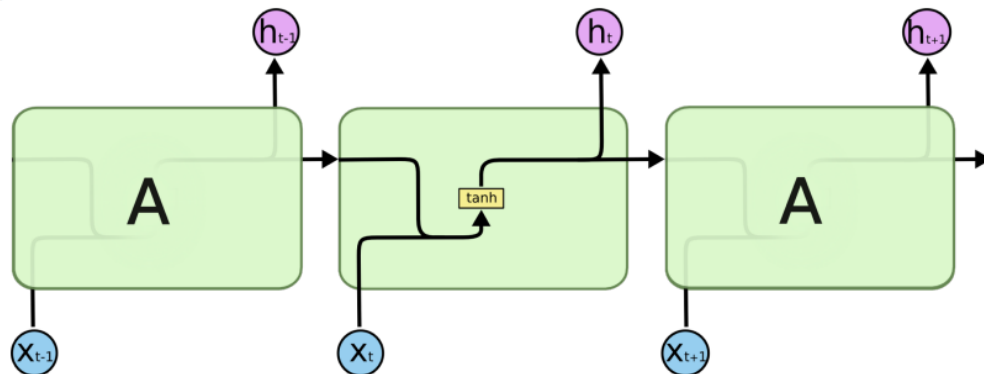
A livello informale, l'output di una LSTM in un dato momento dipende da 3 cose:

1. Dalla *current long-term memory* della rete, detta **cell state**
2. Dall'output al tempo precedente, noto come **previous hidden layer**
3. Dall'input al tempo corrente, noto come **current time step**

Le LSTMs usano una serie di *gates* che controllano come le informazioni di una sequenza di dati temporali vengono gestiti. Ci sono 3 gates: **forget gate**, **input gate**, e **output gate**.

Long-Short Term Memory Network: LSTM

- **Long Short Term Memory networks** – usually just called “**LSTMs**” – are a special kind of RNN, capable of learning **long-term dependencies**. *Hochreiter & Schmidhuber (1997)*
- All recurrent neural networks have the form of a **chain of repeating modules** of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.



Long-Short Term Memory Network: LSTM

- **LSTM** have the same chain like structure except for the repeating module.

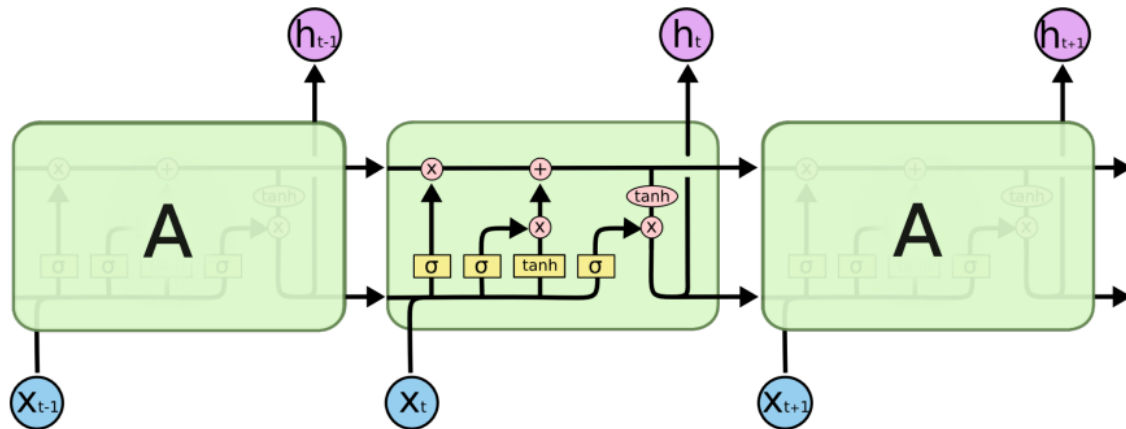
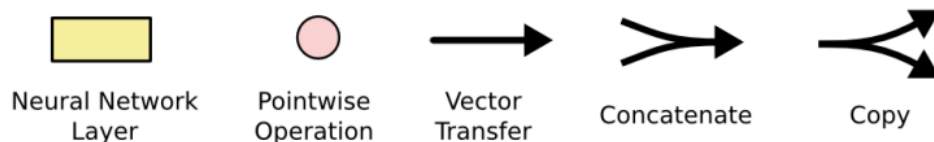
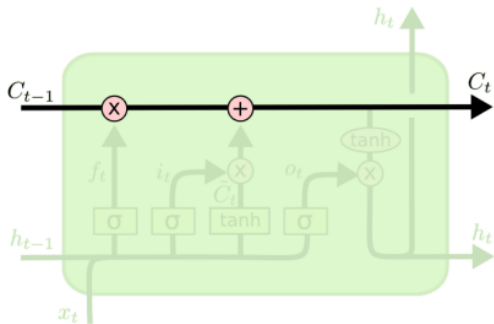


Fig6: The repeating module in a standard RNN contains a single layer [4]

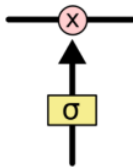


Long-Short Term Memory Network: LSTM

- The core idea behind LSTMs is the **cell state**.



- The LSTM has the ability to **remove** or **add** information to the cell state : thanks to **gates**

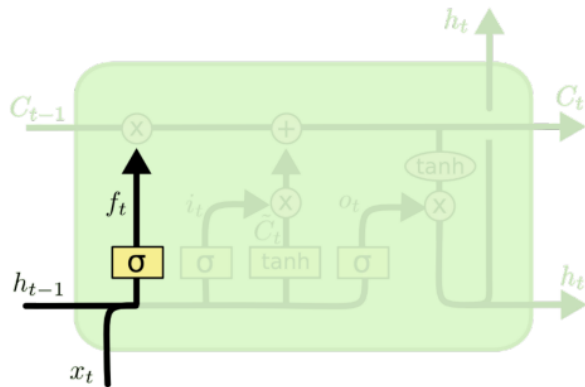


- Gates are composed out of a sigmoid neural net layer and a pointwise multiplication operation

Long-Short Term Memory Network: LSTM

- Step-by-Step LSTM Walk Through

- Step 1:** Decide what information to **throw away** from the cell state, **forget layer**.



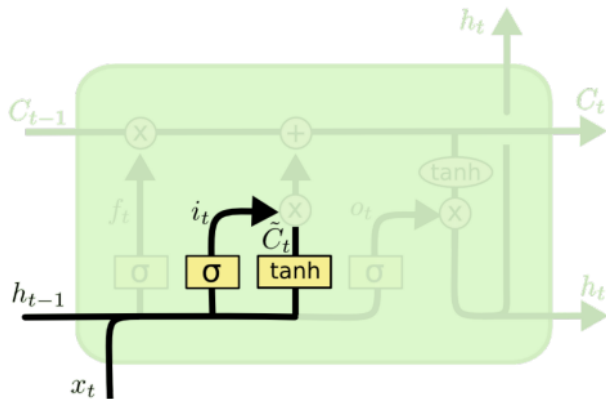
$$f_t = \sigma (W_f \cdot [h_{t-1}, x_t] + b_f)$$

- 1** represents “completely keep this”
- 0** represents “completely get rid of this.”

Long-Short Term Memory Network: LSTM

- Step-by-Step LSTM Walk Through

- Step 2:** Decide what new information we're going to store in the cell state



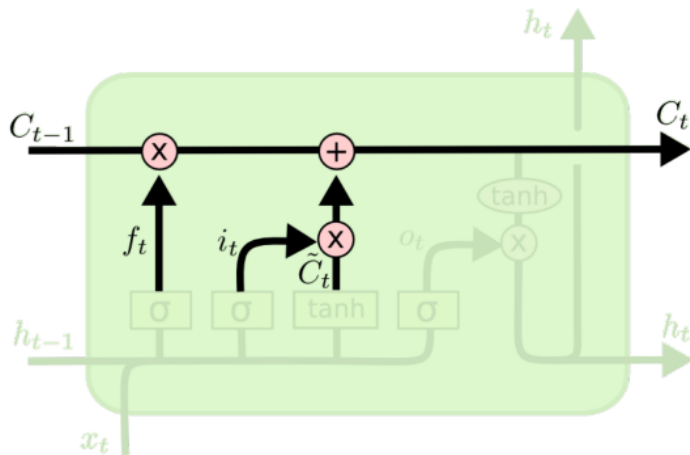
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

- Input gate layer** : decides which values we will update
- Tanh layer** : creates a vector of new candidate values

Long-Short Term Memory Network: LSTM

- Step-by-Step LSTM Walk Through
 - **Step 3:** Update the cell state

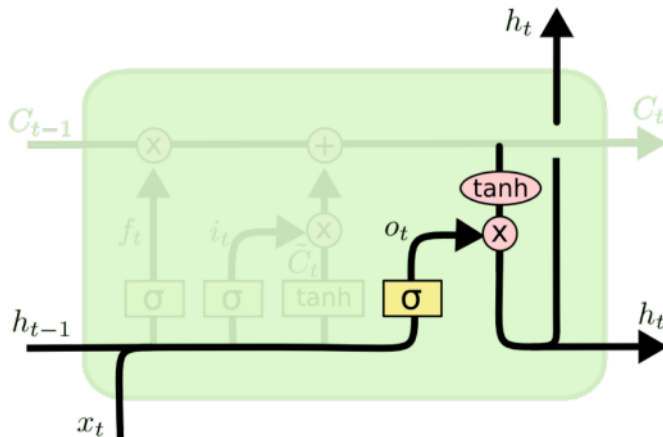


$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

- **Example :** “I grew up in France... I speak fluent *French*.”

Long-Short Term Memory Network: LSTM

- Step-by-Step LSTM Walk Through
 - **Step 4:** Decide what is the output



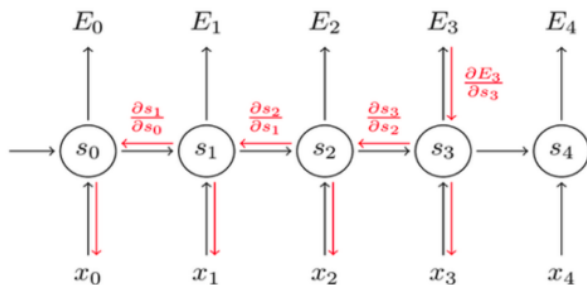
$$o_t = \sigma (W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh (C_t)$$

- **Example :** “I grew up in France... I speak fluent *French*.”

LSTM: Backpropagation Through Time

- **Backpropagation:** Uses partial derivatives and the chain rule to calculate the change for each weight efficiently. Starts with the derivative of the loss function and propagates the calculations backward.
- **Backpropagation Through Time**, or BPTT, is the training algorithm used to update weights in recurrent neural networks like LSTMs.



Il caso di studio: ESAC dataset

ESAC: *Essen Associate Code and Folksong Database* (<http://www.esac-data.org>)

Sviluppato come raccolta di one-part music da European folksong databases

Più di 20000 melodie strumentali soprattutto dalla Germania e Polonia

KERNSCORES Music Collection - <https://kern.humdrum.org/cgi-bin/browse?l=/>



1. si concentra sulle informazioni musicali funzionali, non ortografiche
2. progettato per facilitare le analisi, non la generazione di suoni
3. <https://www.humdrum.org/guide/ch02/>

Music21 score

Score
0.0

meta-
data

(title,
comp,
etc.)
0.0

Part I
0.0

m1a
0.0

clef
0.0

4/4
0.0

q1
0.0

q2
1.0

h1
2.0

m2a
4.0

q3
0.0

h2
1.0

q4
3.0

Part 2
0.0

m1b
0.0

clef
0.0

4/4
0.0

q5
0.0

q6
1.0

h3
2.0

m2b
4.0

q7
0.0

q8
1.0

q9
2.0

q10
3.0

Music21 score: String quartet

The image displays a musical score for a string quartet, featuring four staves: Violin I, Violin II, Viola, and Cello. The music is written in 3/4 time and the key of D major (indicated by two sharps: F# and C#). The score consists of four measures.

- Violin I:** The first measure contains a half note D4, a quarter note E4, and a quarter note F#4. The second measure contains a half note G4 with a trill (tr) and a quarter note F#4. The third measure contains a half note E4 and a quarter note D4. The fourth measure contains a half note C#4 and a quarter rest.
- Violin II:** The first two measures contain whole rests. The third measure contains a half note D4 and a quarter note E4, both marked with a piano (*p*) dynamic. The fourth measure contains a half note F#4 and a quarter note G4, both marked with a piano (*p*) dynamic.
- Viola:** The first measure contains a half note D4 and a quarter note E4, both marked with a piano (*p*) dynamic. The second measure contains a half note F#4 and a quarter note G4. The third measure contains a half note E4 and a quarter note D4. The fourth measure contains a half note C#4 and a quarter rest.
- Cello:** The first measure contains a half note D3 and a quarter note E3, both marked with a piano (*p*) dynamic. The second measure contains a half note F#3 and a quarter note G3. The third measure contains a half note E3 and a quarter note D3. The fourth measure contains a half note C#3 and a quarter note B2.

Score

A musical score for four instruments: Violin I, Violin II, Viola, and Cello. The score is written in 3/4 time and the key of D major (indicated by two sharps: F# and C#). The entire score is enclosed in a red rectangular border.

- Violin I:** The first staff. It begins with a treble clef, a key signature of two sharps, and a 3/4 time signature. The music starts with a half note D4, followed by a dotted half note E4. In the second measure, there is a trill on F#4. The piece concludes with a whole note D4.
- Violin II:** The second staff. It begins with a treble clef, a key signature of two sharps, and a 3/4 time signature. The first two measures contain whole rests. In the third measure, there is a half note D4, followed by a dotted half note E4. The piece concludes with a sixteenth-note ascending scale: F#4, G4, A4, B4, C5, D5.
- Viola:** The third staff. It begins with an alto clef, a key signature of two sharps, and a 3/4 time signature. The music consists of a continuous eighth-note pattern: D4, E4, F#4, G4, A4, B4, C5, D5, repeated throughout the piece.
- Cello:** The fourth staff. It begins with a bass clef, a key signature of two sharps, and a 3/4 time signature. The music consists of a continuous eighth-note pattern: D3, E3, F#3, G3, A3, B3, C4, D4, repeated throughout the piece.

Dynamic markings include *p* (piano) in the first measure of Violin I, the third measure of Violin II, and the first measure of Viola.

Score + Parts + Measures

The image displays a musical score for four instruments: Violin I, Violin II, Viola, and Cello. The score is written in 3/4 time and the key of D major (indicated by two sharps). The music consists of four measures. The Violin I and Viola parts are highlighted with green boxes, while the Violin II and Cello parts are not. The entire score is enclosed in a red box.

Violin I: The first measure contains a half note D4, a quarter note E4, and a quarter note F#4. The second measure contains a half note G4, a quarter note A4, and a quarter note B4. The third measure contains a half note C5, a quarter note B4, and a quarter note A4. The fourth measure contains a half note G4, a quarter note F#4, and a quarter note E4. The part ends with a quarter rest.

Violin II: The first two measures contain whole rests. The third measure contains a half note D4, a quarter note E4, and a quarter note F#4. The fourth measure contains a half note G4, a quarter note A4, and a quarter note B4. The part ends with a quarter rest.

Viola: The first measure contains a half note D4, a quarter note E4, and a quarter note F#4. The second measure contains a half note G4, a quarter note A4, and a quarter note B4. The third measure contains a half note C5, a quarter note B4, and a quarter note A4. The fourth measure contains a half note G4, a quarter note F#4, and a quarter note E4. The part ends with a quarter rest.

Cello: The first measure contains a half note D3, a quarter note E3, and a quarter note F#3. The second measure contains a half note G3, a quarter note A3, and a quarter note B3. The third measure contains a half note C4, a quarter note B3, and a quarter note A3. The fourth measure contains a half note G3, a quarter note F#3, and a quarter note E3. The part ends with a quarter rest.

The dynamic marking *p* (piano) is present in the first measure of Violin I, the third measure of Violin II, and the first measure of Viola.

Score + Parts + Measures

The image displays a musical score for four instruments: Violin I, Violin II, Viola, and Cello. The score is written in 3/4 time and marked *p* (piano). The key signature is one sharp (F#).

The score is enclosed in a red border. The Viola part is highlighted with a green border, and its first four measures are further highlighted with blue boxes. The Cello part is written in the bass clef.

Violin I: The first measure is highlighted with a green box. The notation includes a treble clef, a key signature of one sharp (F#), and a 3/4 time signature. The melody consists of eighth and quarter notes, with a fermata over the final measure.

Violin II: The first measure is highlighted with a green box. The notation includes a treble clef, a key signature of one sharp (F#), and a 3/4 time signature. The instrument is silent for the first two measures, then plays a half note followed by a quarter note, with a fermata over the final measure.

Viola: The first four measures are highlighted with blue boxes. The notation includes a treble clef, a key signature of one sharp (F#), and a 3/4 time signature. The melody consists of eighth and quarter notes, with a fermata over the final measure.

Cello: The first four measures are highlighted with a green box. The notation includes a bass clef, a key signature of one sharp (F#), and a 3/4 time signature. The melody consists of eighth and quarter notes, with a fermata over the final measure.

Score + Parts + Measures + Notes

Violin I

Violin II

Viola

Cello

p

p

p

The image displays a musical score for four instruments: Violin I, Violin II, Viola, and Cello. The score is written in 3/4 time and the key of D major (indicated by two sharps). The Violin I staff is enclosed in a green box. The Viola staff is enclosed in a blue box, and its notes are further highlighted with yellow boxes. The Cello staff is in the bass clef. Dynamics include piano (*p*) and a crescendo. The entire score is framed by a red border.

Grazie

